

Leerwinst en toegevoegde waarde in het voortgezet onderwijs

Eindrapportage



Jaap Roeleveld, Margot Oomens & Alex Coenen

maart 2015

Inhoudsopgave

Voorwoord	5
1 Inleiding	6
1.1 Achtergrond	6
1.2 Doel van de pilots.....	8
1.3 Leeswijzer.....	9
2 Start en verloop van de pilots	10
2.1 Opzet op hoofdlijnen	10
2.2 Start van het project	10
2.3 Begeleiding van de scholen	12
3 Het begeleidende onderzoek	14
3.1 Zicht op opbrengsten van onderwijs.....	14
3.2 Leerwinst en toegevoegde waarde in de pilot.....	18
3.3 Verzamelde gegevens op de scholen	19
3.4 Beschrijvende analyses per school: de schoolrapporten	24
3.5 Slotbeschouwing	31
4 Prestatieniveau en leerwinst vergeleken	32
4.1 Inleiding.....	32
4.2 Prestatieniveau en leerwinst in de pilot	33
4.3 Slotbeschouwing	37
5 Toegevoegde waarde	39
5.1 Inleiding.....	39
5.2 Beginniveau en leerwinst naar afkomst uit APCG-gebied.....	40
5.3 Beginniveau en leerwinst naar ‘voordeel van de twijfel’	44
5.4 Beginniveau en leerwinst naar leerwegondersteuning	46
5.5 Verschillen tussen scholen en cohorten	49
5.6 Slotbeschouwing	51
6 Ervaringen van de scholen	53
6.1 Meerwaarde pilot	53
6.2 Ondersteuning	55
6.3 Onderwerpen pilot.....	56
6.4 Conclusies en adviezen van de scholen.....	57

7	Conclusies en aanbevelingen	60
7.1	Welke mogelijkheden zijn er om de leerwinst van leerlingen in kaart te brengen en te meten?	60
7.2	Welke mogelijkheden zijn er om op basis van de informatie die de school zelf heeft een indicatie van de toegevoegde waarde te verkrijgen en mogelijk tot een gestandaardiseerde methode te komen?.....	61
7.3	Welke rol kan het meten van de leerwinst van leerlingen en het mogelijk meten van een toegevoegde waarde spelen bij de verdere ontwikkeling van het opbrengstgericht werken?	62
7.4	Welke ontwikkelingen binnen de pilots en welke producten uit de pilots kunnen worden ingezet bij het verder ontwikkelen van de opbrengstsystematiek en de ontwikkelingen ten aanzien van het (nieuwe) waarderingskader van de inspectie?	64
	Referenties.....	66
Bijlage 1	Projectgroep en leveranciers.....	69
Bijlage 2	Voorbeelden 'Leerlingeninstroom' uit de schoolrapporten	70

Voorwoord

Leerwinst en toegevoegde waarde zijn begrippen die steeds vaker ter sprake komen als het gaat om opbrengsten van het onderwijs. Maar op welke manier kunnen leerwinst en toegevoegde waarde in het voortgezet onderwijs gemeten worden? En wat kunnen scholen en de inspectie met de uitkomsten van dergelijke metingen? Om dit soort vragen te beantwoorden heeft het Ministerie van OCW pilots uit laten voeren. In deze pilots stonden het ontwikkelen en uitproberen van werkwijzen voor het bepalen van leerwinst en toegevoegde waarde centraal.

Bij de uitvoering van de pilots was een groot aantal partijen betrokken. De directie voortgezet onderwijs van het Ministerie van OCW heeft de pilot mogelijk gemaakt. Zonder haar eindverantwoordelijkheid uit het oog te verliezen, gaf de stuurgroep ons alle ruimte voor de uitvoering van de pilot. De leden van de klankbordgroep hielden ons een spiegel voor zodat we de praktijk op de scholen steeds goed voor ogen hadden. De leveranciers van toets- en schooladministratiesystemen gaven ons een kijkje in de keuken door de mogelijkheden van hun systeem toe te lichten. Ook hebben zij een efficiënte dataverzameling mogelijk gemaakt. De onderzoekers van het Kohnstamm Instituut hebben ervoor gezorgd dat de veelheid aan data geanalyseerd en op een overzichtelijke manier gepresenteerd werd. De adviseurs van Oberon speelden een belangrijke rol bij de ondersteuning van de pilotscholen. Al deze betrokkenen willen we bedanken voor de prettige samenwerking.

Een pilot in het onderwijs kan natuurlijk niet zonder scholen. De pilotscholen hebben hun data ter beschikking gesteld, ze hebben deelgenomen aan gezamenlijke bijeenkomsten en de adviseurs van Oberon werden altijd warm welkom geheten bij hun bezoeken. Veel dank aan de pilotscholen voor hun onmisbare medewerking!

De projectgroep

1 Inleiding

1.1 Achtergrond

Het kerndoel van het beleid is het streven naar de verhoging van de onderwijskwaliteit. De beleidsinstrumenten die daarvoor in de afgelopen jaren zijn ontwikkeld zijn onder meer de wettelijke invoering van referentieniveaus en diverse acties om te stimuleren dat scholen meer opbrengstgericht gaan werken. Opbrengstgericht werken is, in de definitie van de Inspectie van het Onderwijs, “systematisch en doelgericht werken aan het maximaliseren van de prestaties van de leerlingen” (Inspectie van het Onderwijs, 2010). De internationale term is datadriven teaching, hetgeen aangeeft dat het gaat om onderwijs dat mede wordt gestuurd door analyse van gegevens die scholen hebben over leerresultaten van leerlingen (Ledoux e.a., 2009).

Dit beleid is eerder verwoord in de actieplannen ‘Basis voor Presteren’ (primair onderwijs) ‘Beter Presteren’ (voortgezet onderwijs) en ‘Leraar 2020’ (Ministerie OCW, 2011) en meer recent in de Lerarenagenda 2013-2020 (Ministerie OCW, 2013) en de sectorakkoorden met PO en VO (Ministerie OCW & PO-raad, 2014; Ministerie OCW & VO-raad, 2014). Verbetering van leerprestaties, ambitieuze doelen en benutten van talent zijn hierin kernbegrippen. De beleidsstukken bevatten concrete doelen op het gebied van verhogen van de eindopbrengsten van onderwijs, terugdringen van het aantal zwakke scholen, invoering van een predicaat ‘excellente school’ en toename van het aantal scholen dat op een opbrengstgerichte manier werkt, volgens het oordeel van de inspectie.

Nieuwe elementen in deze opeenvolgende plannen zijn een centrale eindtoets voor alle leerlingen in het primair onderwijs, tussentijdse toetsen Nederlands, Engels en rekenen/wiskunde in het voortgezet onderwijs, verplichting voor scholen om gebruik te maken van een leerling- en onderwijsvolgsysteem en de ontwikkeling van een nieuwe maat voor het beoordelen van de leeropbrengsten van een school via het meten van de toegevoegde waarde van de school. Het onderzoek naar de mogelijkheden voor een nieuwe maat moest plaatsvinden in een serie pilots die in het schooljaar 2011-2012 van start zijn gegaan in het primair onderwijs en een jaar later in het voortgezet onderwijs. Het doel is te komen tot een manier van bepalen van toegevoegde waarde die enerzijds functioneel is voor scholen en hen kan helpen bij zelfevaluatie en opbrengstgericht werken en die anderzijds bruikbaar is voor de inspectie bij het beoordelen van de kwaliteit van de leeropbrengsten van scholen, op een manier die zoveel mogelijk objectief en eerlijk is.

De directie voortgezet onderwijs van het ministerie van OCW heeft de inspectie verzocht een projectplan op te stellen voor de pilots in het voortgezet onderwijs (Inspectie van het Onderwijs, 2012). In dit projectplan is kort ingegaan op de termen leerwinst en toegevoegde waarde.¹

¹ Zie ook de uitgebreidere beschouwing over het inzichtelijk maken van onderwijsopbrengsten in paragraaf 3.1.

Omschrijvingen

“De termen ‘toegevoegde waarde’ en ‘leerwinst’ worden vaak naast en door elkaar gebruikt. Ook in het Regeerakkoord en de onderliggende Actieplannen is dat het geval. Het zijn echter twee verschillende begrippen.

*Onder **leerwinst** verstaan we de toename van vaardigheden, kennis en/of competenties van individuele leerlingen of groepen van leerlingen, gedurende (een bepaald deel van) de leerweg. De leerwinst wordt altijd bepaald door minimaal twee meetmomenten. Het verschil tussen die metingen maakt de ontwikkeling van de leerling (of de groep leerlingen) zichtbaar. De metingen moeten eenzelfde schaal omvatten en het moet gaan om genormeerde toetsen.*

*Onder **toegevoegde waarde** verstaan we de bijdrage van de school aan de ontwikkeling (de leerwinst) van leerlingen. Het gaat hierbij dus om dat deel dat met enige zekerheid aan de school kan worden toegeschreven.*

Leerwinst kan nauwkeurig in kaart worden gebracht door te kijken naar twee of meer in de tijd liggende meetmomenten en daar het verschil tussen te meten. Uit de literatuurstudie blijkt dat leerwinstmetingen in het Nederlandse onderwijs het best gebaseerd kunnen worden op individuele metingen van meer dan twee verspreid in de tijd liggende meetmomenten. Toegevoegde waarde is veel lastiger in kaart te brengen en tot een maat te maken: er moet correctie plaatsvinden op een aantal achtergrond- en contextkenmerken (waarbij er nooit honderd procent zekerheid is dat alle correctiefactoren ‘gepakt’ zijn) en er moet vervolgens een schatting gemaakt worden van wat aan de school kan worden toegeschreven. In de pilots gaat het zowel om leerwinst als om de toegevoegde waarde van een school. Op een open wijze zal nader verkend worden hoe de leerwinst van leerlingen in het vo beter recht kan worden gedaan, en hoe dit mogelijk kan leiden tot indicaties van de toegevoegde waarde van een school.

Verskil met pilots primair onderwijs

De pilots Leerwinst en toegevoegde waarde in het primair onderwijs zijn eind 2013 afgerond en in 2014 gerapporteerd (Janssens e.a., 2014). De pilots in het voortgezet onderwijs zijn in 2012 van start gegaan. In deze eindrapportage wordt verslag gedaan van de opbrengsten van het project in het voortgezet onderwijs. Vooraf merken we nog op dat de Ausgangssituatie voor de pilots in het voortgezet onderwijs complex is in vergelijking met die in het basisonderwijs. In het voortgezet onderwijs is, veel minder dan in het basisonderwijs, sprake van het systematisch volgen van de voortgang van leerlingen in een leerlingvolgsysteem (Oomens e.a., 2012; Meijer e.a. 2011).² Daar waar wel gebruik wordt gemaakt van methode-onafhankelijke toetsen, gaat het in het voortgezet onderwijs om uiteenlopende toetsen. Het aanbod van toetsen is in het voortgezet onderwijs breed in vergelijking met het primair onderwijs, waar 85 procent van de scholen gebruik maakt van de Cito-toetsen. Bovendien is er in het voortgezet onderwijs geen sprake van een vergelijkbaar curriculum, maar bieden scholen onderwijs aan op verschillende niveaus, van vmbo basisberoepsgericht tot aan vwo en gymnasium. Daardoor is het moeilijker opbrengsten van scholen onderling te vergelijken.

De bedoeling van de pilots in het vo was daarmee vooral het verkennen van mogelijkheden die scholen voor voortgezet onderwijs hebben om de ontwikkeling van hun leerlingen systematisch te volgen en de uitkomsten daarvan te gebruiken in hun eigen kwaliteitsbeleid en in te zetten bij de ontwikkeling naar meer opbrengstgericht werken. Daarbij hoort ook een verkenning van de mogelijkheden en beperkingen van het gebruik van leerwinst en toegevoegde waarde bij het beoordelen van de opbrengsten van scholen, afdelingen en klassen.

² Overigens worden ook in het basisonderwijs de mogelijkheden van leerlingvolgsystemen lang niet altijd optimaal gebruikt (Meijer, Ledoux & Elshof, 2011).

1.2 Doel van de pilots

Het project had tot doel gedurende twee schooljaren informatie te verzamelen over de mogelijkheden om tot een bruikbare, valide maat voor leerwinst te komen, evenals informatie over de mogelijkheden voor een gestandaardiseerde methode voor het bepalen van de toegevoegde waarde van een school voor voortgezet onderwijs. Om tot die maten en methoden te komen, is aangesloten bij de huidige schoolpraktijk, zoals de pilotscholen die laten zien. De werkwijze om leerwinst en toegevoegde waarde te bepalen moet (statistisch) valide en betrouwbaar zijn, maar ook recht doen aan de inspanningen van de school en rekening houden met verschillen tussen scholen in het aangeboden niveau van onderwijs, de gebruikte toetsen en hun schoolpopulatie.

Belangrijk kenmerk van het project was dat er werd uitgegaan van de bestaande schoolpraktijk en dat er gebruik gemaakt zou worden van gegevens zoals die op dat moment reeds bij de scholen aanwezig waren. In het voortgezet onderwijs staat het gebruik van leerlingvolgsystemen, waarmee de vooruitgang van leerlingen systematisch kan worden bijgehouden, nog in de kinderschoenen. En waar dergelijke systemen wel worden gehanteerd worden zij zelden ingezet voor opbrengstgericht werken (Oomens e.a., 2012). Overigens leert de ervaring in het basisonderwijs dat, ook wanneer er wel op grote schaal gewerkt wordt met leerlingvolgsystemen, er vaak gebrek aan deskundigheid op school is wat betreft het benutten van de mogelijkheden die deze systemen bieden (zie bijvoorbeeld Meijer e.a., 2011). Een van de opbrengsten van de pilots zou dan ook moeten zijn de scholen handvatten te geven om de ontwikkeling van hun eigen leerlingen systematisch te volgen. Voor schoolmanagers, schoolbestuurders en docenten zouden daarmee instrumenten ter beschikking moeten komen om het eigen onderwijs te beoordelen en te verbeteren.

In de richting van de Inspectie dienden de pilots informatie op te leveren over hoe leerwinst en toegevoegde waarde van afdelingen en scholen voor voortgezet onderwijs kunnen worden bepaald en mogelijk een rol kunnen spelen bij het beoordelen van de opbrengsten van scholen. Daarbij gaat het enerzijds om een betrouwbare en valide manier van beoordelen. Maar anderzijds dient zo'n beoordeling ook transparant en begrijpelijk voor de scholen en besturen te zijn. En de benodigde gegevens moeten zonder te veel administratieve last voor de scholen algemeen beschikbaar zijn.

Voor de pilot werden in het eerder genoemde projectplan de volgende onderzoeksvragen geformuleerd (Inspectie van het Onderwijs, 2012, pag. 4):

1. Welke mogelijkheden zijn er om de leerwinst van leerlingen in kaart te brengen en te meten?
2. Welke mogelijkheden zijn er om op basis van de informatie die de school zelf heeft een indicatie van de toegevoegde waarde te verkrijgen en mogelijk tot een gestandaardiseerde methode te komen?
3. Welke rol kan het meten van de leerwinst van leerlingen en het mogelijk meten van een toegevoegde waarde spelen bij de verdere ontwikkeling van het opbrengstgericht werken?
4. Welke ontwikkelingen binnen de pilots en welke producten uit de pilots kunnen worden ingezet bij het verder ontwikkelen van de opbrengstsystematiek en de ontwikkelingen ten aanzien van het (nieuwe) waarderingskader van de inspectie?

In het voorliggende eindrapport doen wij verslag van de verrichte werkzaamheden binnen de pilots en komen we tot conclusies en aanbevelingen met betrekking tot deze vragen voor het project.

1.3 Leeswijzer

In hoofdstuk 2 worden de start en het verloop van de pilots beschreven. Daarbij komen de verwachtingen van scholen bij de start van het project en de verdere begeleiding van de scholen gedurende het project aan de orde. Hoofdstuk 3 doet verslag van het begeleidende onderzoek bij de pilots. Na een inleidende beschouwing over het beoordelen van onderwijsopbrengsten wordt nader ingegaan op de bij de scholen verzamelde gegevens, de verwerking daarvan en de terugrapportages van beschrijvende analyses naar de scholen in de vorm van schoolrapporten. In hoofdstuk 4 maken we een vergelijking gemaakt tussen prestatieniveau en leerwinst. Daarbij wordt geïllustreerd dat verschillende manieren om naar de opbrengsten van scholen te kijken ook tot verschillende waarderingen van die opbrengsten kunnen leiden. Hoofdstuk 5 gaat nader in op toegevoegde waarde. Daarbij wordt aangetoond dat het van belang is om rekening te houden met verschillen tussen scholen in de (achtergrond)kenmerken van hun leerlingen. In hoofdstuk 6 staan de ervaringen van de scholen met betrekking tot hun deelname aan de pilots centraal. Daarbij komen de schoolrapporten en de begeleiding vanuit het project aan de orde. Ook worden een aantal conclusies en adviezen vanuit de scholen gerapporteerd. We sluiten het rapport in hoofdstuk 7 af met conclusies en aanbevelingen, waarbij we antwoorden geven op de hiervoor genoemde onderzoeksvragen voor de pilots.

2 Start en verloop van de pilots

2.1 Opzet op hoofdlijnen

De pilots Leerwinst en toegevoegde waarde in het voortgezet onderwijs hebben gelopen vanaf het najaar van 2012 tot eind 2014. De pilots zijn een initiatief van het ministerie van OCW, dat de Inspectie van het Onderwijs gevraagd heeft de projectleiding te verzorgen. Bij de uitvoering van de pilot waren twee organisaties betrokken:

- Het Kohnstamm Instituut was verantwoordelijk voor de opzet en uitvoering van het wetenschappelijk onderzoek voor de pilot (onder andere de data-analyses en opstellen schoolrapportages);
- Oberon was verantwoordelijk voor de ondersteuning van de projectgroep, ondersteuning bij de dataverzameling op de scholen, bezoeken van de deelnemende scholen in het kader van terugkoppeling van de resultaten en het organiseren en/of verzorgen van inhoudelijke ondersteuning op maat. Elke school had tijdens de pilot in principe een vaste contactpersoon/ adviseur vanuit Oberon.

Voor de pilots is een projectorganisatie samengesteld, waarbij de volgende groepen worden onderscheiden: stuurgroep, projectgroep en klankbordgroep. De stuurgroep, bestaande uit vertegenwoordigers van OCW en de Inspectie van het Onderwijs, was eindverantwoordelijk voor de pilots. De projectgroep, met daarin vertegenwoordigers van OCW, de inspectie, het Kohnstamm Instituut en Oberon, was verantwoordelijk voor de uitvoering van de pilots.³ De voortgang van de pilots en de tussen- en eindopbrengsten zijn door de projectgroep voorgelegd aan de klankbordgroep. Daarin zaten vertegenwoordigers van de VO-raad, de AOC Raad, Diataal, Cito en een schoolleider. Om goed zicht te krijgen op de mogelijkheden van de diverse toetsystemen hebben leden van de projectgroep gesprekken gevoerd met leveranciers van deze systemen.⁴

Belangrijk uitgangspunt van de pilot was om zoveel mogelijk aan te sluiten bij de bestaande gegevens/praktijk op de scholen. Daarom werd ook gestreefd naar een zo efficiënt mogelijke manier van aanleveren. Oberon heeft deze dataverzameling gecoördineerd en uitgevoerd. Ander uitgangspunt binnen de pilot was dat als er nieuwe gegevens verzameld moesten worden dit geen extra kosten met zich mee zou brengen voor de scholen. Voor het meten van motivatie en burgerschapscompetenties is daarom gekozen voor de vrij beschikbare vragenlijsten uit het landelijke cohortonderzoek COOL⁵⁻¹⁸ (zie paragraaf 3.2).

2.2 Start van het project

Werving en aanmelding

Via een oproep op de websites van de inspectie en de VO-raad zijn scholen geattendeerd op de pilot en konden zij zich aanmelden. Daarnaast heeft de inspectie een aantal scholen gevraagd om te deel te nemen als naar aanleiding van een inspectiebezoek duidelijk was dat de school veel interesse had in het onderwerp van de pilot. Uiteindelijk hebben 27 vo-scholen zich definitief aangemeld voor de pilot. Deze scholen liggen verspreid over Nederland en bieden onderwijs aan van vmbo-basis tot en met gymnasium. Per school stonden een of twee afdelingen centraal in de pilot. Gedurende de pilot hebben twee scholen zich terug getrokken omdat zij binnen de school andere prioriteiten stelden dan verdere deelname aan de pilot.

³ Zie bijlage 1 voor een overzicht van de leden van de projectgroep.

⁴ Zie bijlage 1 voor een overzicht van de betreffende organisaties.

Gezamenlijke startbijeenkomst

Het project is gestart met een gezamenlijke bijeenkomst op 2 november 2012 in Utrecht. Tijdens deze bijeenkomst konden de pilotscholen met elkaar kennismaken en met vertegenwoordigers van de inspectie en OCW, de onderzoekers van het Kohnstamm Instituut en de adviseurs van Oberon. Daarnaast werden de scholen geïnformeerd over de verdere opzet van de pilot en werden vragen van de scholen geïnventariseerd.

Intakegesprek

Na de gezamenlijke startbijeenkomst is door de begeleiders van Oberon op alle afzonderlijke scholen een intakegesprek gehouden. Het doel van deze gesprekken was om meer zicht te krijgen op de (toets)gegevens die scholen in pilot beschikbaar hadden en wilden gebruiken voor leerwinst en toegevoegde waarde en op de wijze waarop en de mate waarin scholen opbrengstgericht werken. Ook werd aan de scholen gevraagd welke behoeften zij hadden aan verdere ondersteuning rondom leerwinst en toegevoegde waarde. De gesprekken bevestigden dat de scholen gebruik maken van diverse toetssystemen (Cito, Diataal, TOA, Route VO, Deviant) en schooladministratiesystemen (SOM, Magister). Bij de toetssystemen wordt het Cito Volgsysteem voortgezet onderwijs het meest gebruikt (13 pilotscholen in het eerste jaar en 15 in het tweede jaar). Zes scholen in het eerste jaar en vier in het tweede gebruiken Diataal. TOA wordt door twee pilotscholen gebruikt en de overige toetssystemen worden door elk een pilotschool. Een school heeft geen toetsstelsel in gebruik. Uit de intakegesprekken bleek dat scholen verschillen in de wijze waarop zij opbrengstgericht werken en in de ervaring die ze daarmee hebben.

De verwachtingen van scholen van de pilot waren wisselend. Een deel van de scholen wilde met behulp van de pilot ervoor zorgen dat de Inspectie van het Onderwijs bij de beoordeling meer rekening houdt met de achtergrondkenmerken van hun leerlingen. Zij gaven bijvoorbeeld aan een lastige leerlingpopulatie te hebben waar de inspectiebeoordeling voor hun gevoel te weinig rekening mee houdt. Andere scholen hoopten met behulp van de pilot en de daarin ontwikkelde instrumenten op het gebied van leerwinst en toegevoegde waarde meer zicht te krijgen op de tussentijdse ontwikkeling van leerlingen, de ontwikkeling van leerlingen in bepaalde schooltypen of bijvoorbeeld op het onderwijsrendement in de onder- of bovenbouw. Enkele scholen wilden meer grip krijgen op de afstroom van leerlingen of docenten meer bewust maken van het belang van opbrengstgericht werken. Ze hoopten dat de pilot handvatten zou geven om het onderwijs te verbeteren. Daarnaast waren er scholen met andere verwachtingen, zoals zicht krijgen op de rol van kleinschaligheid of de relatie met dubbele adviezen of overadvisering.

Sommige scholen gaven aan niet alleen te willen kijken naar de cognitieve ontwikkeling van hun leerlingen, maar ook op zoek te zijn naar manieren om de niet-cognitieve ontwikkeling meetbaar te maken. Daarom is dit ook dit ook meegenomen in de pilot. Bij een aantal scholen bleek verwarring te zijn over het begrip toegevoegde waarde. Zij dachten dat toegevoegde waarde betrekking had op de zaken die de school leerlingen leert naast de cognitieve vakken. In de pilot gaat het bij toegevoegde waarde echter om de bijdrage van de school aan de leerwinst van leerlingen, ongeacht of deze leerwinst betrekking heeft op de cognitieve ontwikkeling of op andere aspecten.

Tot slot is in de intakegesprekken aan de scholen gevraagd aan welke ondersteuning zij behoefte hadden. Sommige scholen hadden vooral behoefte aan het bespreken van de schoolrapportage van de eerste pilotmeting. Anderen gaven aan behoefte te hebben aan een workshop waarin dieper werd ingegaan op wat je als school kunt met gegevens uit diverse toetsen of rondom opbrengstgericht werken in bredere zin. Er waren ook scholen die meer wilden weten over de reden van afstroom van leerlingen in de onderbouw of juist van het succes van hun leerlingen in het vervolgonderwijs.

De bevindingen uit de intakegesprekken zijn gebruikt om:

- de dataverzameling af te stemmen op de door scholen gebruikte systemen en beschikbare gegevens;
- de keuze te bepalen voor instrumenten voor het meten van de niet-cognitieve ontwikkeling;
- het ondersteuningsaanbod zoveel mogelijk af te stemmen op de wensen van de scholen (ook gebaseerd op de bevindingen uit de besprekingen van de eerste schoolrapportages).

2.3 Begeleiding van de scholen

Tijdens de pilot werden de scholen begeleid door een adviseur van Oberon. De begeleiding bestond naast het verzorgen van de dataverzameling uit het bespreken van rapportages die door het Kohnstamm Instituut werden opgesteld per school, het inrichten van een projectwebsite, gezamenlijke bijeenkomsten voor de scholen over de pilot en ondersteuning op maat per school (indien gewenst).

Dataverzameling

Voor de pilots is gebruik gemaakt van gegevens afkomstig uit drie verschillende bronnen:

- schooladministratiesystemen (achtergrondgegevens van leerlingen en eindexamengegevens);
- toetssystemen;
- digitale vragenlijsten motivatie en burgerschap.

Omdat niet alle scholen binnen de pilot gebruik maakten van dezelfde gegevens, heeft elke school een instructie op maat ontvangen met daarin welke gegevens geleverd moesten worden en op welke manier. Voor vragen over de gegevenslevering konden scholen terecht bij de helpdesk van Oberon. Alle scholen hebben op basis van een gerichte instructie achtergrondgegevens van de leerlingen vanuit hun schooladministratiesysteem aangeleverd. Voor de twee meest gebruikte systemen (SOM en Magister) heeft Oberon deze instructie opgesteld in overleg met de leveranciers van de systemen. Voor de andere systemen heeft Oberon samen met de betreffende scholen gekeken wat de meest efficiënte manier was voor het aanleveren van de gegevens.

Beschikbare gegevens uit methode-onafhankelijke toetssystemen zijn via de leveranciers zelf opgevraagd na toestemming van de scholen (Cito, Diataal en TOA). De twee scholen die andere toetsen gebruikten, hebben de toetsgegevens van hun leerlingen zelf geleverd.

De vragenlijsten uit het landelijke cohortonderzoek COOL⁵⁻¹⁸ zijn door Oberon gedigitaliseerd. Scholen die (een van) deze vragenlijsten gebruikten binnen de pilot, hebben van Oberon de link naar de vragenlijst en de inloggegevens ontvangen. De scholen waren zelf verantwoordelijk voor het laten invullen van de vragenlijst door de leerlingen.

Website

Bij de start van de pilots heeft Oberon een website ontworpen en ingericht voor alle betrokkenen, pilotscholen, stuurgroep, projectgroep en klankbordgroep. Op deze website was informatie opgenomen over de achtergrond van de pilots, de pilotscholen, het onderzoek, de ondersteuning van de scholen en de data en verslagen van overleggen en bijeenkomsten. Gedurende de eerste periode voorzag de website in de informatiebehoefte van betrokkenen. In de loop van de pilot bleek echter dat zij een voorkeur hadden om rechtstreeks per e-mail geïnformeerd te worden. Bovendien waren ook niet bij de pilot betrokkenen geïnteresseerd in informatie over de pilots. Daarom na het eerste pilotjaar besloten de projectwebsite niet langer te onderhouden, maar de

belangrijkste informatie (kort en bondig) op te nemen op de website van de inspectie⁵ en de betrokkenen voortaan per e-mail te informeren.

Terugkoppeling schoolrapportages

De scholen hebben twee maal een terugkoppeling gekregen van de resultaten van de pilot in de vorm van een schoolrapportage. De rapportages werden besproken op de scholen door een adviseur van Oberon. De gesprekken werden in een aantal gevallen bijgewoond door een medewerker van OCW of de inspectie. In de besprekingen werden de resultaten aan de scholen inhoudelijk toegelicht, konden de scholen vragen stellen over de rapportage en hier feedback op geven. Vanuit de school was bij de bespreking meestal een lid van de directie aanwezig en vaak ook een of meer afdelings- of teamleiders.

Gezamenlijke bijeenkomsten

Naast de eerder genoemde startbijeenkomst zijn er twee andere gezamenlijke bijeenkomsten georganiseerd voor de pilotscholen: een tussentijdse bijeenkomst en een slotbijeenkomst. Bij de bijeenkomsten waren niet alleen de pilotscholen, maar ook de projectgroep, de onderzoekers van het Kohnstamm Instituut en de adviseurs van Oberon aanwezig. Voor de slotbijeenkomst waren daarnaast ook andere geïnteresseerde scholen uitgenodigd.

De tussentijdse bijeenkomst vond plaats na afloop van het eerste pilotjaar. Tijdens deze bijeenkomst zijn de pilotscholen geïnformeerd over de voortgang van de pilot en over de tussentijdse resultaten en werd vooruitgeblikt naar de opzet en planning van het tweede pilotjaar. Daarnaast hebben de pilotscholen in subgroepen kennis en ervaringen uitgewisseld. In de slotbijeenkomst zijn de ervaringen uitgewisseld met en tussen de pilotscholen en geïnteresseerden vanuit het onderwijsveld (andere vo-scholen, toetsontwikkelaars, vertegenwoordigers van OCW en inspectie) en is het project voor de pilotscholen afgesloten.

Ondersteuning op maat

De pilotscholen konden ook ondersteuning op maat krijgen. De invulling van de ondersteuning werd bepaald in overleg met de scholen, de projectgroep en de onderzoekers. In totaal hebben 12 scholen een of meerdere vormen van ondersteuning op maat gekregen. De overige scholen hebben op eigen verzoek hier geen gebruik van gemaakt. In de meeste gevallen ging het om een workshop voor docenten of schoolleiding, meestal op het gebied van opbrengstgericht werken. Daarnaast zijn voor een aantal scholen aanvullende analyses uitgevoerd door het Kohnstamm Instituut. Op twee scholen is een bijeenkomst georganiseerd over afstroom en bij een andere school is geholpen bij het leggen van contacten met scholen die gebruik maken van hetzelfde toetsstelsel (Deviant). Ook is op een school samen met de directie gekeken naar welke mogelijkheden de op de school beschikbare gegevens bieden voor eigen analyses door de school.

⁵ <http://www.onderwijsinspectie.nl/onderwijs/Voortgezet+onderwijs/leerwinst-en-toegevoegde-waarde>

3 Het begeleidende onderzoek

In de pilots ging het om de vraag hoe de opbrengsten van onderwijs inzichtelijk kunnen worden gemaakt; en meer in het bijzonder om de vraag hoe de opbrengsten van afdelingen en scholen kunnen worden beoordeeld. Rondom de beoordelingen van de opbrengsten van onderwijs zijn verschillende benaderingen mogelijk. We geven hier in paragraaf 3.1 eerst een algemeen overzicht daarvan, in hoofdstuk 3.2 en 3.3 gaan we nader in op hoe een en ander in de pilots aan de orde is gekomen. In paragraaf 3.4 staan de zogenaamde schoolrapporten centraal. In paragraaf 3.5 geven we een slotbeschouwing.

3.1 Zicht op opbrengsten van onderwijs

In onderstaand schema staat een samenvattend overzicht van de verschillende benaderingen die mogelijk zijn rondom de beoordelingen van de opbrengsten van onderwijs.⁶ In de rest van deze paragraaf lichten we deze benaderingen verder toe.

Schema 3.1 Benaderingen van opbrengsten van onderwijs

Niveau	Uitgangspunt	Type vraagstelling
Individueel	1. prestaties	Wat kan/kent een leerling op een bepaald moment? Wat is zijn of haar niveau op dat moment?
	2. leerwinst	Hoeveel heeft een leerling geleerd in een bepaalde periode? Wat is zijn of haar ontwikkeling
School	1. prestaties	Wat is het gemiddelde niveau van leerlingen op deze school op dit moment? Is dat hoger of lager dan bij andere scholen?
	2. <i>value added</i> *	Zijn de gemiddelde prestaties van leerlingen op deze school hoger of lager dan verwacht kon worden op basis van een aantal kenmerken van de leerlingen?
	3. leerwinst	Hoeveel hebben de leerlingen op deze school gemiddeld geleerd in een bepaalde periode? Is dat meer of minder dan bij andere scholen?
	4. <i>value added</i> *	Hebben de leerlingen op deze school gemiddeld meer of minder geleerd dan verwacht kon worden op basis van een aantal kenmerken van de leerlingen?
Stelsel	1. prestaties	Wat is het gemiddelde niveau van Nederlandse leerlingen op een bepaald meetmoment (bv. einde basisonderwijs; op 15-jarige leeftijd; enz.)? Is dat hoger of lager dan in voorafgaande jaren? Hoger of lager dan andere OECD-landen?
	2. <i>value added</i> *	Is het gemiddelde niveau van Nederlandse leerlingen op een bepaald meetmoment hoger of lager dan verwacht kon worden op basis van een aantal kenmerken van de leerlingen (nu en in de voorafgaande jaren)?
	3. leerwinst	Hoeveel leren Nederlandse leerlingen in een bepaalde fase van het onderwijs (bijv. gedurende de basisschool; gedurende de onderbouw vo; enz.)? Is dat meer of minder dan in voorafgaande cohorten?
	4. <i>value added</i> *	Is de leerwinst van Nederlandse leerlingen in een bepaalde fase van het onderwijs meer of minder dan verwacht kon worden op basis van een aantal kenmerken van die leerlingen (nu en in de voorgaande cohorten)?

* Toegevoegde waarde.

⁶ We maken hier gebruik van onze eerdere beschouwing in Roeleveld, Van der Veen en Ledoux (2008).

Onderwijsopbrengsten bij leerlingen (niveau individueel)

Uitgangspunt bij de beoordeling van de opbrengsten van onderwijs zijn de resultaten van **individuele leerlingen**. Daar zijn in principe twee varianten van.

1. De *prestatie* van een leerling op een bepaald moment (een 'bruto' onderwijsresultaat). We kunnen daarbij bijvoorbeeld denken aan:
 - de score op de eindtoets basisonderwijs;
 - het wel of niet behalen van een referentieniveau voor taal en rekenen;
 - in het voortgezet onderwijs: de behaalde onderwijspositie (volgens de zogenaamde leerjarenladder) na 3 jaar voortgezet onderwijs;
 - het cijfer voor het eindexamen in bepaalde (kern)vakken;
 - enzovoort.
2. De *leerwinst* van een leerling op een bepaald moment (een 'netto' onderwijsresultaat). Hierbij wordt uitgegaan van een beginmeting en wordt gekeken hoeveel de leerling vooruit is gegaan tussen de beginmeting en de prestatie aan het eind. We kunnen hierbij denken aan:
 - het verschil in vaardigheidsscore voor taal en rekenen in groep 8 met de score op dezelfde schaal in groep 4;
 - het verschil tussen toetsscores aan het begin en het eind van de eerste klas voortgezet onderwijs;
 - het verschil tussen de behaalde onderwijspositie na 3 jaar voortgezet onderwijs (volgens de leerjarenladder) en het advies voor vo van de basisschool (vergelijk het onderbouwrendement in de inspectiebeoordeling);
 - enzovoort.

Bij de tweede variant wordt er dus rekening mee gehouden dat sommige leerlingen het onderwijs (of preciezer: een bepaalde fase van het onderwijs) binnenkomen met meer 'bagage' dan andere leerlingen. Beide varianten van deze individuele resultaten kunnen vervolgens gebruikt worden ter beoordeling van de opbrengsten van scholen én als indicator van het onderwijsstelsel als geheel.

Onderwijsopbrengsten bij scholen

Bij het beoordelen van de opbrengsten van scholen gaat het er meestal om te bepalen welke scholen het duidelijk 'beter doen' dan gemiddeld en welke scholen het duidelijk 'slechter doen' dan gemiddeld. Het gaat dus om een relatieve waardering. Net als bij de opbrengsten op individueel niveau kunnen ook opbrengsten op het niveau van de school bepaald worden op basis van prestaties en op basis van leerwinst. Deze 'output-maten' geven een soort absoluut beeld van de school, zonder dat rekening gehouden wordt met de (soms zeer grote) verschillen tussen scholen in hun leerlingbevolking. Voor een 'eerlijke' vergelijking van scholen zou gekeken moeten worden naar hun toegevoegde waarde. Er worden 'value added models' toegepast, waarin langs statistische weg wél rekening kan worden gehouden met die verschillen in leerlingpopulatie.⁷ In feite wordt er dan, op basis van een aantal leerlingkenmerken (zoals opleiding van de ouders, etniciteit, verblijfsduur in Nederland, thuistaal, enz.) een voorspelling gedaan over wat er van een bepaalde leerling verwacht mag worden. Daarmee zijn er vier mogelijkheden voor het in beeld brengen van onderwijsopbrengsten op schoolniveau.

⁷ Zie bijvoorbeeld OECD (2008).
Recente overzichten van de (internationale) literatuur en toepassingen van zulke modellen zijn te vinden in Timmermans (2012) en Janssens, Rekers-Mombarg & Lacor (2014).

1. Uitgaand van de (individuele) *prestaties* van de leerlingen kan een samenvattende maat per school bepaald worden, bijvoorbeeld:
 - het gemiddelde op de eindtoets basisonderwijs;
 - het percentage leerlingen dat het minimum- of het streefniveau voor taal of rekenen behaalt;
 - de gemiddelde onderwijspositie na 3 jaar voortgezet onderwijs; of
 - het gemiddelde eindexamencijfer voor bepaalde (kern)vakken;
 - enz.
2. Om een eerlijkere vergelijking tussen scholen mogelijk te maken kunnen, zoals gezegd, *prestaties* in beeld worden gebracht met '*value added models*'. Bij de vergelijking van de opbrengsten van scholen wordt dan gekeken naar vragen als:
 - is de gemiddelde score op de eindtoets basisonderwijs hoger of lager dan voorspeld kan worden op basis van een aantal kenmerken van de leerlingen in groep 8?
 - is het percentage leerlingen dat het minimum- of het streefniveau voor taal of rekenen behaalt hoger of lager dan verwacht kon worden?
 - is de gemiddelde onderwijspositie na 3 jaar voortgezet onderwijs hoger of lager dan voorspeld?
 - is het gemiddelde eindexamencijfer voor bepaalde (kern)vakken hoger of lager dan verwacht kon worden?
 - enz.
3. Uitgaand van de (individuele) *leerwinst* van de leerlingen kan ook een maat per school bepaald worden. Bijvoorbeeld:
 - de gemiddelde leerwinst voor taal en rekenen tussen groep 4 en groep 8?
 - de gemiddelde leerwinst tussen begin en eind van de eerste klas in het vo?
 - de gemiddelde vooruitgang op de leerjarenladder tussen het advies, waarmee de leerlingen het vo binnenkomen en hun onderwijspositie na 3 jaar voortgezet onderwijs?
 - enz.
4. Zoals gezegd is leerwinst een 'netto' onderwijsresultaat, waarbij er rekening wordt gehouden met het feit dat de ene leerling bij binnenkomst al meer kan dan de andere. Maar ook hier kunnen '*value added models*' toegepast worden om tot een eerlijker vergelijking van scholen te komen. Want ook de ontwikkeling van een leerling tussen de 'beginmeting' en de 'eindmeting' wordt beïnvloed door meer of minder stimulerende factoren van buiten de school, waardoor het op de ene school makkelijker is om tot een bepaalde *leerwinst* te komen dan op de andere school. Bij de beoordeling van de opbrengsten van scholen wordt dan gekeken naar vragen als:
 - is de gemiddelde leerwinst voor taal en rekenen hoger of lager dan voorspeld kan worden op basis van een aantal kenmerken van de leerlingen?
 - is de gemiddelde leerwinst in de eerste klas van het vo hoger of lager dan verwacht mocht worden van de betreffende leerlingen?
 - is de gemiddelde vooruitgang op de leerjarenladder tussen advies en onderwijspositie na 3 jaar (onderbouwrendement) hoger of lager dan verwacht kon worden voor deze leerlingpopulatie?
 - enz.

De voorspellingen en verwachtingen bij '*value added models*' worden bepaald door vergelijking met de resultaten van leerlingen op een groot aantal scholen in een representatieve steekproef.

Opbrengsten van het stelsel

De belangrijkste functie van een indicator voor het onderwijs op stelselniveau is dat deze de mogelijkheid geeft om na te gaan hoe de historische ontwikkeling is: is er sprake van voor- of achteruitgang over de loop der jaren. Een andere functie kan zijn dat de indicator een rol speelt in internationale vergelijkingen (PISA, TIMMS, enz.). Ook op stelselniveau kan een indicator gebaseerd zijn op ('bruto') prestatie of op ('netto') leerwinst; en ook op

stelselniveau kan er met rechtstreekse gemiddelden gewerkt worden of met modellen voor toegevoegde waarde.

1. Uitgaand van de (individuele) *prestaties* van de leerlingen kan een samenvattende maat voor het hele onderwijsstelsel bepaald worden. Bijvoorbeeld:

- het gemiddelde op de eindtoets basisonderwijs;
- het percentage leerlingen dat het minimum- of het streefniveau voor taal of rekenen behaalt;
- de gemiddelde onderwijspositie na 3 jaar voortgezet onderwijs;
- het gemiddeld eindexamencijfer voor bepaalde (kern)vakken;
- enz.

Hiermee krijgen we een maat voor de opbrengsten van het Nederlandse onderwijsstelsel op een bepaald moment. Door vergelijking van een aantal opeenvolgende momenten kan een beeld verkregen worden van de historische ontwikkelingen (“stijgt of daalt het percentage leerlingen dat het minimumniveau voor rekenen behaalt?”). Voor sommige (beleids-)vragen kan dit een uitstekende indicator zijn. Bijvoorbeeld als we willen weten of het Nederlandse basisonderwijs er in slaagt om 75 procent van de leerlingen op een minimumniveau voor rekenen te krijgen, ongeacht de beginsituatie of de (sociale en etnische) herkomst van die leerlingen.

2. Voor andere vragen is het wel nuttig om rekening te houden met de herkomst van de leerlingen. Wanneer bijvoorbeeld over een (lange) reeks van jaren de *prestaties* in groep 8 vergeleken worden, dan kan het nodig zijn om rekening te houden met verschuivingen in de leerlingbevolking over die reeks van jaren (zoals bv. een toename van het aandeel leerlingen uit etnische groepen; een stijging van het ouderlijk onderwijsniveau; enz.). Hiervoor kunnen ook ‘*value added models*’ worden toegepast; nu niet zozeer voor een ‘eerlijke’ vergelijking tussen scholen met een verschillende leerlingpopulatie, maar voor een ‘eerlijke’ vergelijking van leerlingpopulaties door de tijd.

3. Om een goed beeld te krijgen van wat het onderwijsstelsel presteert kan ook uitgegaan worden van de *leerwinst* van leerlingen. Er kan dan bijvoorbeeld, net als bij de scholen hiervoor, maar nu op stelselniveau, gekeken worden naar:

- de gemiddelde leerwinst voor taal en rekenen tussen groep 4 en groep 8;
- de gemiddelde leerwinst in de eerste klas van het vo;
- de gemiddelde vooruitgang op de leerjarenladder tussen het advies, waarmee de leerlingen het vo binnenkomen en hun onderwijspositie na 3 jaar voortgezet onderwijs;
- enz.

4. Zoals gezegd wordt met *leerwinst* in elk geval rekening gehouden met het feit dat leerlingen bij het ingaan van een bepaalde fase van het onderwijs verschillen in hun aanvangsniveau. Maar ook hun verdere groei na dat aanvangsniveau wordt beïnvloed door buitenschoolse factoren. Daarom kan het nuttig zijn ook hier met complexere ‘*value added models*’ te werken. Een extra voordeel daarvan is dat er ook zicht kan komen op de ontwikkeling van specifieke groepen leerlingen (bijvoorbeeld leerlingen van bepaalde etnische herkomst), wat bij kan dragen tot een ‘rijkere’ stelselindicator dan bijvoorbeeld een enkel percentage dat een referentieniveau behaalt.

Samenvattend

Met als uitgangspunt de resultaten van individuele leerlingen (‘bruto’ prestaties of ‘netto’ leerwinst) kan de rechtstreekse output van een onderwijsinstelling, een onderwijssector, of een heel onderwijsstelsel bepaald worden of kan desgewenst de toegevoegde waarde van de instelling, sector of stelsel geschat worden met ‘*value added models*’. Daarmee kunnen verschillende typen vragen rond onderwijsopbrengsten beantwoord worden, zoals die in het schema aan het begin van deze paragraaf zijn weergegeven.

Er is wetenschappelijke overeenstemming over de gedachte dat de kwaliteit van scholen niet zomaar afgemeten kan worden aan de ruwe, ongecorrigeerde (eind)prestaties van hun leerlingen, maar dat er rekening

moet worden gehouden met het beginniveau van leerlingen en/of met achtergrondkenmerken van leerlingen die hun onderwijssucces mede beïnvloeden (Roeleveld 2003, Onderwijsraad, 2003, Wijnstra e.a., 2003; OECD, 2008; Roeleveld e.a., 2008; Timmermans, 2012; Dijkstra & Janssens, 2012; Rekers-Mombarg e.a., 2014). Daarom is in deze pilots gezocht naar manieren om, op basis van bij de school zelf aanwezige gegevens, leerwinst en toegevoegde waarde in beeld te krijgen.

3.2 Leerwinst en toegevoegde waarde in de pilot

In het project Pilots leerwinst en toegevoegde waarde in het voortgezet onderwijs is gezocht naar bruikbare indicatoren voor de kwaliteit van een school, of van een afdeling in een school. De aandacht gaat daarbij met name uit naar indicatoren die betrekking hebben op leeropbrengsten bij leerlingen. We gebruiken hier de term leeropbrengsten in ruime zin, ongeacht of deze op het cognitieve vlak liggen (zoals leeropbrengsten binnen de afzonderlijke schoolvakken) of op het sociaal-emotionele domein (zoals motivatie, zelfvertrouwen of burgerschapscompetenties).

In het project zijn twee typen indicatoren onderscheiden: indicatoren op basis van leerwinst en indicatoren op basis van toegevoegde waarde.⁸

Leerwinst

Indicatoren op basis van leerwinst zijn afgeleid van de leerwinst die de leerlingen van de school (of afdeling) in een nader aan te geven periode hebben geboekt. Idealiter is het voor de bepaling van de leerwinst nodig dat leerlingen een begintoets en een eindtoets afleggen. Aangenomen dat de toetsscores op één schaal kunnen worden gebracht, komt het verschil tussen de beginmeting en de eindmeting in dat geval overeen met de leerwinst.

Als de betreffende toetsen landelijk genormeerd zijn kan voor het beoordelen van de behaalde leerwinst (gemiddeld per klas of per school) een vergelijking gemaakt worden met de leerwinst die landelijk behaald wordt. Op die manier wordt de leerwinst dus vergeleken met landelijke gemiddelden: er wordt vergeleken met 'doorsnee-leerlingen'.

Toegevoegde waarde

Maar scholen trekken een zeer verschillend leerlingenpubliek. Met een beginmeting weten we iets over het niveau waarmee de leerlingen de school binnen komen. Maar daarnaast spelen er ook andere factoren mee, die bevorderend of belemmerend kunnen zijn voor de (cognitieve en sociale) ontwikkeling van de leerlingen. Dat kan een individuele problematiek betreffen, maar ook de sociale achtergrond van leerlingen speelt mee. Bij sommige leerlingen is het moeilijker vooruitgang te boeken dan bij andere leerlingen. Hoe beter dat in kaart kan worden gebracht, des te beter kunnen we ook de bijdrage van de school zelf aan de ontwikkeling van leerlingen inschatten.

Daarmee komen we bij de indicatoren op basis van toegevoegde waarde. Deze worden geoperationaliseerd als dat deel van de leerwinst dat als de unieke bijdrage van de school kan worden beschouwd. In dit type indicatoren wordt een correctie toegepast voor verschillen tussen leerlingen waar scholen geen grip op hebben, zoals intelligentie of leerpotentieel of de (meer of minder gunstige) omstandigheden thuis. Via statistische modellen, waarin scholen met andere scholen worden vergeleken, worden toetsscores of

⁸ In de internationale literatuur wordt overigens leerwinst (d.w.z. rekening houden met het beginniveau van leerlingen) vaak al aangeduid met 'value added', de Engelse term voor toegevoegde waarde.

examencijfers van leerlingen ontleed in afzonderlijke componenten. Alleen de component die geen statistische samenhang vertoont met aanvangsverschillen tussen leerlingen (zoals intelligentie, omstandigheden thuis, voorkennis) wordt beschouwd als een schatting van de toegevoegde waarde. Met andere woorden, het onverklaarde deel in de kennisgroei van leerlingen wordt beschouwd als de unieke bijdrage van de school.

Verskil leerwinst en toegevoegde waarde

Er is een belangrijk verschil tussen de beide typen indicatoren. Leerwinstindicatoren hebben tot op zekere hoogte een absoluut karakter: schattingen van de leerwinst voor één school zijn onafhankelijk van schattingen van de leerwinst voor andere scholen. Maar in deze absolute zin boeken (vrijwel) alle leerlingen wel leerwinst in een bepaald deel van hun onderwijsloopbaan. Beoordeling van de behaalde leerwinst in termen van voldoende of onvoldoende gebeurt dan door vergelijking met landelijke gemiddelden, afkomstig uit de normeringsonderzoeken van de toetsontwikkelaars. De onafhankelijk van andere scholen te schatten leerwinst wordt vervolgens dus wel relatief beoordeeld, als gezegd door vergelijking met 'doorsnee-leerlingen'.

Indicatoren op basis van toegevoegde waarde hebben veel duidelijker een relatief karakter. Bij het bepalen van toegevoegde waarde wordt rekening gehouden met verschillen tussen scholen in hun leerlingenpubliek. Bij sommige leerlingen is het makkelijker om vooruitgang te boeken dan bij andere leerlingen. Bij leerlingen met 'ongunstige' achtergrondkenmerken zullen de verwachtingen omtrent de te behalen leerwinst lager zijn dan bij andere leerlingen. En een school met veel leerlingen met 'ongunstige' achtergrondkenmerken zal meer moeite moeten doen om de landelijke normen, gebaseerd op 'doorsnee-leerlingen', te behalen dan een school met veel leerlingen met juist 'gunstige' achtergrondkenmerken. Voor de bepaling van welke verwachtingen we bij welke leerlingen mogen hebben, zijn geen rechtstreekse landelijke vergelijkingsgegevens voorhanden. Deze moeten gehaald worden uit onderzoek waarin de resultaten van scholen met verschillende leerlingpopulaties onderling worden vergeleken. De schattingen van de toegevoegde waarde voor één school zijn daarmee afhankelijk van de vorderingen die leerlingen op andere scholen boeken. Pas door opbrengsten van scholen met verschillende leerlingpopulaties met elkaar te vergelijken kan bepaald worden of een school relatief veel of juist weinig vooruitgang boekt met de leerlingen.

Omdat scholen over het algemeen alleen over eigen vorderingengegevens beschikken – en niet over die van andere scholen – zijn indicatoren op basis van leerwinst voor hen veel makkelijker te bepalen dan indicatoren op basis van de toegevoegde waarde. Daar komt nog bij dat voor het schatten van het laatste type indicatoren statistische expertise onontbeerlijk is.

3.3 Verzamelde gegevens op de scholen

Zoals eerder opgemerkt werd er in de pilots uitgegaan van gegevens, zoals die op de scholen zelf aanwezig zijn. Het idee hierachter is dat scholen zelf, met hun eigen gegevens, leerwinst en toegevoegde waarde moeten kunnen gaan bepalen en beoordelen.

In het voortgezet onderwijs worden onderwijsresultaten, cijfers en toetsscores vooral gebruikt om te determineren: waar is de leerling op zijn/haar plek (op- en afstroom), kan de leerling overgaan of moet hij/zij een jaar over doen? Maar informatie over onderwijsresultaten zou nog beter benut kunnen worden door meer systematisch de ontwikkeling van leerlingen, klassen en afdelingen te volgen. Hoe staat de leerling er voor bij instroom op de school; hoe ontwikkelt hij/zij zich vervolgens gedurende de schooljaren? Hoe gaan de onderwijsresultaten van een klas vooruit (leerwinst)?

Omdat scholen onderling verschillen in het soort leerlingen dat instroomt in de school is het ook van belang zicht te krijgen op kenmerken van die instroom. De instromende leerlingen vormen het ‘materiaal’ waarmee de scholen moeten gaan werken (Bosker, 2012). Bij de ene leerling kan immers ‘makkelijker’ vooruitgang en leerwinst geboekt worden dan bij de andere leerling.

In het project zijn verschillende soorten gegevens van de scholen verkregen:

- kenmerken van leerlingen uit de schooladministraties;
- scores op een aantal verschillende toetsen, die gebruikt kunnen worden om de cognitieve ontwikkeling van leerlingen in beeld te brengen;
- eindexamencijfers;
- gegevens uit de vragenlijsten Motivatie en Burgerschapscompetenties.

Niet alle scholen hebben met al deze gegevens meegedaan met de pilot. Per school kwamen verschillende soorten gegevens beschikbaar, met als voornaamste onderscheid of het om cognitieve resultaten (toetsen, eindexamencijfers) ging of om sociaal-emotionele opbrengsten (motivatie, burgerschap). Per afzonderlijke school zijn dan ook steeds wisselende analyses gedaan die naar de scholen zijn teruggerapporteerd. In de volgende paragraaf geven we een algemeen beeld van de uitkomsten van deze schoolanalyses. Hier gaan we eerst nader in op de verschillende soorten verzamelde gegevens.

Gegevens uit de schooladministraties

Van de scholen zijn gegevens verkregen van de leerlingen, die in de onderzochte schooljaren (2012/13 en 2013/14) op de school zaten. Algemene gegevens zijn leerlingnamen, geboortedata, geslacht en het (interne) stamnummer van de leerling.⁹ Daarnaast zijn gegevens verkregen over het leerjaar en het type onderwijs (vmbo-basis, vmbo-kader, vmbo-gl/tl, havo, vwo; en combinaties daarvan in brugklassen). Zoals gezegd zijn dit gegevens vanuit deze twee schooljaren. Als een leerling nu, bijvoorbeeld, in leerjaar 4 in een havo klas zit weten we niet precies op welk niveau die leerling gestart is in het voortgezet onderwijs. Dat kan in een brugklas geweest zijn met havo, maar er kan ook sprake zijn geweest van eerdere op- of afstroom. In de analyses en de schoolrapporten zijn de leerlingen ingedeeld naar het type onderwijs in het betreffende schooljaar.

In de ideale situatie wordt voor de analyse van leerwinsten en van toegevoegde waarde uitgegaan van een cohortbenadering: alle leerlingen, die in een bepaald leerjaar instromen in de school (een cohort) doen mee met de toetsen en worden door de jaren heen longitudinaal gevolgd. Dan kan ook op- en afstroom en uitval (of verhuizing) van leerlingen verdisconteerd worden. In deze pilot moeten we echter uitgaan van de beperkte dwarsdoorsnedes van schooljaren 2012/13 en 2013/14.

Verder is aan de scholen gevraagd om een aantal nadere kenmerken van de leerlingen. Met deze gegevens wilden we een indicatie krijgen van de uitgangssituatie, waarmee leerlingen in de school instromen. Idealiter wilden we beschikken over een schatting van het cognitieve niveau en eerdere schoolprestaties van leerlingen. En daarnaast over indicaties rond een meer of minder stimulerend thuisklimaat en eventuele bijzondere omstandigheden. Voor het aanvangsniveau van de leerlingen is gevraagd naar de score op de Cito-eindtoets basisonderwijs of de NIO en het advies voor voortgezet onderwijs dat de leerling heeft gekregen. Gebleken is dat dit niet op alle scholen systematisch geregistreerd wordt. Het advies, dat voor alle leerlingen verstrekt wordt, is soms maar deels opgenomen in de administratie. En zowel de Cito-eindtoets als de NIO worden niet

⁹ Namen en geboortedata waren nodig om per individuele leerling achtergrondgegevens en toetsresultaten uit verschillende schooljaren te kunnen koppelen. Daarbij zijn de privacy-regels voor wetenschappelijk onderzoek gehanteerd. Deze identificerende gegevens worden na afloop van het onderzoek vernietigd.

op alle basisscholen afgenomen. En ook wanneer deze wel zijn afgenomen wordt de score niet altijd systematisch geregistreerd in de schooladministratie van de school voor voortgezet onderwijs.

Een andere indicatie voor het aanvangsniveau is het al dan niet vertraagd zijn in de eerdere loopbaan in het basisonderwijs. Deze is voor het eerste leerjaar te bepalen uit de geboortedatum van de leerlingen, die voor bijna alle leerlingen is opgenomen in de administratiebestanden. Bij de hogere leerjaren is doorgaans niet meer scherp te bepalen of de vertraging al bij instroom bestond of dat die is opgelopen op de pilotschool zelf.

Wat betreft indicaties van het thuisklimaat zouden we het liefst beschikken over het opleidingsniveau en de etnische herkomst van de ouders. Uit tal van onderzoeken (zie bijvoorbeeld Onderwijsraad, 2013) blijken deze kenmerken samen te hangen met de onderwijsresultaten van de kinderen. Vanwege de gewichtenregeling wordt in het basisonderwijs doorgaans geregistreerd of de ouders lage opleidingen hebben, hoewel het onderscheid tussen middelbare en hogere opleidingen ook daar vaak ontbreekt. In het voortgezet onderwijs is de opleiding van de ouders echter onbekend. Als grove benadering van de thuissituatie kan daar alleen gewerkt worden met het al dan niet wonen in een zogenaamd APCG: een armoedeprobleemcumulatiegebied (SCP/CBS, 2007). Vanuit de postcodes van leerlingen in de schooladministraties kan deze indicator voor vrijwel alle leerlingen bepaald worden; incidenteel ontbrak zo'n postcode in de aangeleverde gegevens.

Of er sprake is van bijzondere omstandigheden bij de leerlingen, wordt zeer wisselend geregistreerd door de scholen. Onduidelijk is ook of dit systematisch is gebeurd. In het algemeen is wel bekend of er sprake is van leerwegondersteuning (lwoo).

Wanneer scholen de ontwikkeling van leerlingen systematisch willen gaan bijhouden en ook willen nagaan hoe verschillende groepen leerlingen zich ontwikkelen, dan zou een eerste stap moeten zijn om goed te registreren wat er bekend is over de 'uitgangssituatie' van de startende leerlingen: wat is bekend over hun eerdere loopbaan (advies, Cito en of NIO, vertraging) en wat voor speciale omstandigheden zijn er.

Toetsscores

In de pilot is gebruik gemaakt van een aantal methode-onafhankelijke toetsen, die betrekking hebben op de domeinen Nederlandse taal, rekenen/wiskunde en Engels. Om de vooruitgang van leerlingen op deze domeinen scherp in beeld te krijgen zijn er prestatiemetingen van leerlingen nodig op tenminste twee tijdstippen. Leerwinst wordt pas echt meetbaar als de toetsscores op die twee meetmomenten ook op eenzelfde schaal zijn uit te drukken. Van de bij de scholen gebruikte toetsen voldoen aan deze eisen:

- de Cito-toetsen uit het Volgsysteem voortgezet onderwijs (VVO, voorheen VAS);
- de Diataal toetsen;
- de toetsen van Route VO (voorheen Toets 234VO).

De twee andere toetsen, die scholen in de pilot gebruiken (TOA en Deviant), worden gebruikt om de mate van beheersing van de referentieniveaus voor taal en rekenen te bepalen. Er worden geen vaardigheidsscores, die op eenzelfde schaal kunnen worden geplaatst, mee bepaald.

Verder moeten de toetsen ook valide en betrouwbaar zijn en voorzien zijn van landelijke normen. Een van de eisen die we in dit verband aan de toetsen kunnen stellen is dat ze als voldoende of goed zijn beoordeeld door de COTAN (Commissie Testaangelegenheden Nederland van het Nederlands Instituut van Psychologen). Van de in de pilot gebruikte toetsen zijn alleen de Cito-toetsen en de Route VO toetsen door de COTAN beoordeeld en op de meeste criteria als voldoende of goed beoordeeld. Bij deze toetsen bestaan ook landelijke normtabellen. Diataal is (nog) niet beoordeeld door de COTAN, maar wel wetenschappelijk onderbouwd (Hacquebord e.a.,

2005). Diataal geeft normtabellen die jaarlijks worden aangepast. Van TOA en Deviant is geen landelijk normeringsonderzoek bekend.

Voor de meeste scholen zijn, na toestemming van de scholen hiervoor, de toetsscores centraal verkregen bij de toetsaanbieders (Cito, Diataal, TOA). Doorgaans konden ook historische gegevens, van toetsafnames uit eerdere jaren, worden verstrekt. Het per leerling koppelen van deze scores aan gegevens uit de schooladministraties heeft in de praktijk de nodige problemen opgeleverd. De verkregen gegevens bestaan uit een groot aantal losse bestanden (per afnamejaar en per toets). Bij Diataal en TOA is daarbij veelal ook het (unieke) stamnummer opgenomen, waaronder de leerlingen ook bekend zijn in het administratiebestand. Bij Cito zijn de leerlingen alleen identificeerbaar aan de hand van hun naam, geboortedatum en geslacht. Vanwege spelfouten, verwisseling van voor- en achternaam, ontbrekende geboortedata, enzovoort, is de koppeling daar niet volledig geslaagd. Voor statistische doeleinden hoeft dit niet zo'n probleem te zijn, maar wel als men de ontwikkeling van een individuele leerling wil kunnen volgen.

Wanneer de scholen zelf de ontwikkeling van de toetsscores van leerlingen systematisch willen gaan volgen zou het aan te bevelen zijn dat scholen de uitkomsten van de verschillende toetsen ook zelf in hun leerlingvolgsystemen gaan registreren. Of dat gebruik wordt gemaakt van, bijvoorbeeld, de mogelijkheid om in het Cito-portal ook het stamnummer van de leerlingen op te nemen. Maar het zou nog beter zijn als er een automatische koppeling tot stand zou kunnen komen tussen het volgsysteem en de toetsregistratiesystemen.

Tot slot merken we nog op dat al de gebruikte toetsen bepaalde cognitieve vaardigheden meten. In de curricula van de onderzochte scholen komen natuurlijk ook allerlei andere vaardigheden aan de orde, van bijvoorbeeld de beroepsgerichte vakken in het vmbo tot aan de klassieke vorming en talen op het gymnasium. Voor zulke vakken zijn over het algemeen echter geen landelijk genormeerde instrumenten voorhanden, die het mogelijk maken om opbrengsten van scholen (in termen van leerwinst en toegevoegde waarde) met elkaar te vergelijken.

Eindexamencijfers

Er zijn enkele pilotscholen die ook hun opbrengsten in termen van eindexamengegevens nader willen analyseren, bij voorkeur in verband gebracht met (toets)gegevens van leerlingen bij instroom in de school. Het aantrekkelijke daarvan is dat de schoolloopbaan als geheel, vanaf het begin (start leerjaar 1) tot en met het eindexamen, onderzocht wordt. Tegelijk brengt zo'n langere periode van tenminste vier leerjaren ook een aantal problemen met zich mee: er is tussentijds sprake van zittenblijven, op- en afstroom en in- en uitstroom op de school. De bij de scholen verkregen eindexamencijfers bieden niet goed zicht op dat hele tussenliggende proces. Bovendien kan met eindexamencijfers pas op het eind van het onderwijsproces een beoordeling van de onderwijsopbrengsten plaats vinden. Tussentijdse bijstelling van dat proces voor de betreffende leerlingen is dan niet meer mogelijk. Eindexamencijfers passen daarom eerder in een 'accountability' benadering (summatieve evaluatie) dan in een benadering gericht op onderwijsverbetering (formatieve evaluatie).

Vragenlijsten Motivatie

Deze vragenlijst wordt gebruikt in het landelijke cohortonderzoek COOL⁵⁻¹⁸ (Driessen e.a., 2012; Zijlsing e.a., 2012).¹⁰ Uit de vragenlijst kunnen scores op de hieronder omschreven schalen worden bepaald. De schalen gaan over doelen voor leren. Zo werken sommige leerlingen graag hard, omdat ze het beter willen doen dan andere leerlingen (Prestatie), anderen doen hun best op school, omdat ze graag complimentjes en beloningen ontvangen (Extrinsiek).

¹⁰ Zie ook www.cool5-18.nl.

Bekwaming (mastery) betreft de mate waarin leerlingen gericht zijn op het ontwikkelen van hun eigen vaardigheden en het verbeteren daarvan. Items zijn bijvoorbeeld: 'Ik zie graag dat mijn schoolwerk steeds beter wordt' en 'Ik doe goed mijn best op school omdat ik mijn schoolwerk interessant vind'.

Prestatie (performance) gaat over je best doen op school om beter te presteren in vergelijking met andere leerlingen en het sociaal gezien verkrijgen van aanzien. Items zijn bijvoorbeeld: 'Ik ben alleen maar tevreden als ik een van de besten van de klas ben' en 'Ik doe mijn best op school zodat ik de leider van een groepje word'

Sociaal (social). Leerlingen die hier hoog op scoren, waarderen het sociale aspect van leren: ze voelen zich betrokken bij andere kinderen en waarderen de omgang met andere kinderen. Voorbeelditems zijn 'Ik help graag andere kinderen goede cijfers te halen op school' en 'Ik werk op school liever samen met andere kinderen dan alleen'.

Extrinsiek (extrinsic) gaat over harder willen werken voor school als je complimentjes en beloningen ontvangt. Items zijn bijvoorbeeld: 'Ik krijg graag complimentjes voor goed schoolwerk' en 'Als je op school 'bonuspunten' zou kunnen krijgen, zou ik beter mijn best doen'.

Met **taakmotivatie** wordt bedoeld de mate waarin leerlingen gericht zijn op het ontwikkelen van hun vaardigheden, op leren en het begrijpen van schoolwerk. Leerlingen zijn daarbij gericht op hun eigen prestaties en hoe die zich ontwikkelen en niet op de vergelijking met andere leerlingen. Items zijn bijvoorbeeld: 'Ik ben tevreden als ik op school iets heb geleerd dat ik begrijp' en 'Als ik op school iets niet meteen snap, ga ik er juist extra mijn best voor doen'.

Met **cognitief zelfvertrouwen** (self-efficacy) wordt bedoeld de mate waarin leerlingen vertrouwen in zichzelf hebben dat zij kunnen slagen op school. Hier zijn items bijvoorbeeld: 'Ik weet zeker dat dit jaar alles op school me wel zal lukken' en 'Ik kan bijna alles op school, als ik het maar blijf proberen'.

De eerste vier schalen zijn een vertaalde versie van de Inventory of School Motivation van McInerney & Ali (2006). Taakmotivatie is gemeten door gebruik te maken van een schaal ontwikkeld door Seegers, Van Putten & De Brabander (2002). Bij cognitief zelfvertrouwen is gebruik gemaakt van de schaal 'self-efficacy' die afkomstig is van de 'Patterns of Adaptive Learning Survey (PALS)' (Midgley e.a., 2000). Alle genoemde schalen laten ook op de gegevens in dit pilot onderzoek een goede betrouwbaarheid zien (Cronbach's alpha van .78 tot .88).

Vragenlijsten Burgerschapscompetenties

Ook deze vragenlijst wordt gebruikt in het landelijke cohortonderzoek COOL⁵⁻¹⁸ (Driessen e.a., 2012; Zijsling e.a., 2012).¹¹ Het meetinstrument burgerschap (Ten Dam e.a., 2010) is ontwikkeld als antwoord op toenemende aandacht voor burgerschapsvorming in het onderwijs. In het instrument is het begrip competentie uiteengelegd in vier componenten: kennis, reflectie, vaardigheden en attituden. Het begrip burgerschap is geoperationaliseerd aan de hand van vier centrale sociale taken:

- democratisch handelen: het aanvaarden van en bijdragen aan een democratische samenleving;
- maatschappelijke verantwoordelijkheid: medeverantwoordelijkheid nemen voor de leefgemeenschappen waartoe men behoort;

¹¹ Zie ook www.cool5-18.nl.

- omgaan met conflicten: betreft (lichte) conflictsituaties of belangentegenstellingen waarbij de jongere zelf 'partij' is;
- omgaan met verschillen: betreft sociale, culturele, religieuze en uiterlijke verschillen.

Bij elk van deze vier centrale taken worden de componenten kennis, reflectie, vaardigheden en attitudes met een aantal items gemeten.

Schema 3.3 Inhoud vragenlijst Burgerschapscompetenties (vier schalen)

28 reflectie-items: deze gaan over reflecteren, nadenken over. De voorgelegde vraag is: 'hoe vaak denk jij na over...', bijvoorbeeld '...of er naar leerlingen wordt geluisterd op jouw school?' en '.....hoe het komt dat mensen van verschillende afkomst weinig met elkaar omgaan?'

15 vaardigheden-items: hier moet de leerling aangeven hoe goed hij/zij zichzelf ergens in vindt. De voorgelegde vraag is 'hoe goed ben jij in...', bijvoorbeeld '...indenken hoe een ander zich voelt en daarmee rekening houden?' en '...je taalgebruik aanpassen aan degene met wie je spreekt?'

24 attitude-items: deze gaan over iets vinden (menen), willen, tot iets bereid zijn. De voorgelegde vraag is 'hoe goed past een uitspraak bij jou?'. Uitspraken zijn bijvoorbeeld: 'Als iemand in de klas het ergens mee oneens is, moet hij/zij de kans krijgen om dat uit te leggen' en 'Ik vind het leuk om iets te weten van verschillende soorten geloof'.

27 kennis-items: deze gaan over weten, begrijpen, inzicht hebben in wat je het beste kunt doen in een specifieke situatie. Bij de kennis-items kiezen leerlingen het beste antwoordalternatief uit bij een vraag, bijvoorbeeld: 'Alle kinderen hebben recht: a) op zakgeld, b) om te kiezen bij wie je woont, c) op onderwijs'. In dit geval is optie c het goede antwoord.

In de analyses en schoolrapporten worden de vier hoofdschalen Attitude, Vaardigheid, Reflectie en Kennis gebruikt. De score op Vaardigheid is dan bijvoorbeeld de samenvatting van alle vaardigheden met betrekking tot democratisch handelen, maatschappelijke verantwoordelijkheid, omgaan met conflicten én omgaan met verschillen. We hebben de betrouwbaarheid van de vier schalen ook op de gegevens in dit pilot onderzoek nagegaan en er blijkt steeds sprake te zijn van een goede betrouwbaarheid (Cronbach's alpha steeds groter dan .85).

Beide vragenlijsten zijn door leerlingen van een aantal pilotscholen online ingevuld. Identificatie van de leerlingen en koppeling van deze gegevens aan de administratiebestanden van de scholen moest gebeuren aan de hand van leerlingnamen en geboortedata. Net als bij de toetsscores bleek dat niet altijd te lukken, bijvoorbeeld doordat leerlingen geen naam hebben ingevuld.

3.4 Beschrijvende analyses per school: de schoolrapporten

In de schoolrapporten van schooljaren 2012/13 en 2013/14 zijn naar de scholen beschrijvende gegevens teruggekoppeld. We geven hierna een globale beschrijving van deze analyses per school. Wat er feitelijk in het rapport van een specifieke school stond, was natuurlijk afhankelijk van het soort gegevens (type toets, motivatie en burgerschap, eindexamencijfers), waarmee de school deelnam aan de pilot. Er is wel voor alle scholen zoveel mogelijk dezelfde opzet gevolgd: steeds een beschrijving van de kenmerken van hun leerlingeninstroom, gevolgd door een aantal uitkomsten van het onderwijs (cognitief en/of sociaal-emotioneel) op de school.

Kenmerken van de leerlingeninstroom

De scholen in de pilot verschillen in hun leerlingpopulatie. Dat hangt uiteraard samen met welke afdeling van de school meedoet in het onderzoek. Maar ook scholen die met een zelfde soort afdeling meedoen, kunnen onderling nog sterk verschillen, bijvoorbeeld naar het aandeel leerlingen met lwoo, het aandeel leerlingen uit een armoedeprobleemcumulatiegebied, het aandeel leerlingen dat met een hoger of juist een lager advies aan de afdeling begint.

In de schoolrapporten werden gegevens van de instroom in beide onderzochte schooljaren vergeleken met landelijke cijfers. Deze laatste cijfers zijn afkomstig uit landelijke registerbestanden (BRON, gebaseerd op het onderwijsnummer) en voor deze pilots bewerkt door de Inspectie van het Onderwijs. Omdat de samenstelling van (brede) brugklassen nogal heterogeen is, werden ook cijfers gegeven van het derde leerjaar, als de nadere uitsplitsing naar onderwijstype, ook in de landelijke gegevens, voltooid is. Voor zover bekend uit de op de school aanwezige gegevens ging het om de volgende achtergrondkenmerken van de leerlingen:

- percentage meisjes;
- percentage leerlingen uit een armoedeprobleemcumulatiegebied (APCG);
- percentage leerlingen met indicatie voor leerwegondersteuning (lwoo);
- percentage leerlingen dat bij instroom vertraagd is.

Daarnaast werden een aantal gegevens gepresenteerd die een beeld kunnen geven van het niveau van de leerlingen bij de afsluiting van het basisonderwijs:

- percentage leerlingen met een advies dat hoger of lager was dan de gekozen afdeling;
- percentage leerlingen met een score op de Cito-eindtoets basisonderwijs die hoger of lager is dan de gekozen afdeling;¹²
- percentage leerlingen met een score op de NIO (of incidenteel een andere IQ-test) die hoger of lager is dan de gekozen afdeling;¹³

Bij de overzichten¹⁴ van achtergrondkenmerken van de leerlingen kon de school haar eigen instroom vergelijken met de landelijk cijfers van vergelijkbare scholen. Daaruit kan dan bijvoorbeeld blijken of er op de school relatief veel of juist weinig leerlingen zitten met leerwegondersteuning. Of dat de school veel meer of juist veel minder leerlingen afkomstig uit een APCG-gebied heeft dan scholen met dezelfde afdeling(en) elders in het land.

De kenmerken van de afsluiting van het basisonderwijs geven de school een beeld van het aandeel leerlingen, dat met het 'voordeel van de twijfel' is ingestroomd: leerlingen waarvan het advies of de score op de eindtoets eigenlijk te laag was voor het niveau van voortgezet onderwijs waar zij zijn ingestroomd. Ook hier kon de school nagaan of zij relatief veel of juist weinig van zulke leerlingen heeft toegelaten, in vergelijking met soortgelijke scholen elders in het land.

De overzichten van hun eigen instroom waren daarmee bedoeld om de school haar eigen verhaal achter de cijfers te kunnen laten onderbouwen: is onze leerlingenpopulatie nu 'moeilijker' of 'makkelijker' dan die van andere scholen, die hetzelfde type onderwijs aanbieden.

¹² Hoger of lager is gebaseerd op: *Betekenis van de standaardscore op de Cito-toets 2012*, Arnhem: Cito.

¹³ Hoger of lager is gebaseerd op: Van Dijk & Telligen, 2004, *De NIO en schoolkeuze*. Amsterdam, Boom. Tabel 5.6.

¹⁴ In de Bijlage staan voorbeelden van de tabellen die rond de leerlingeninstroom aan de scholen zijn gepresenteerd. Zie voor een geanonimiseerd schoolrapport de website van de Inspectie van het Onderwijs:
<http://www.onderwijsinspectie.nl/onderwijs/Voortgezet+onderwijs/leerwinst-en-toegevoegde-waarde>.

Behaalde toetsresultaten

Op de pilotscholen werden verschillende toetsen gebruikt, die soms sinds kort en soms al langere tijd werden afgenomen bij de leerlingen. Het gaat om de volgende toetsen:

- Cito-toetsen uit het Volgsysteem voortgezet onderwijs (VVO, voorheen bekend als VAS);
- Diataal toetsen;
- toetsen van Route VO (voorheen 234VO)
- TOA-toetsen;
- Deviant toetsen.

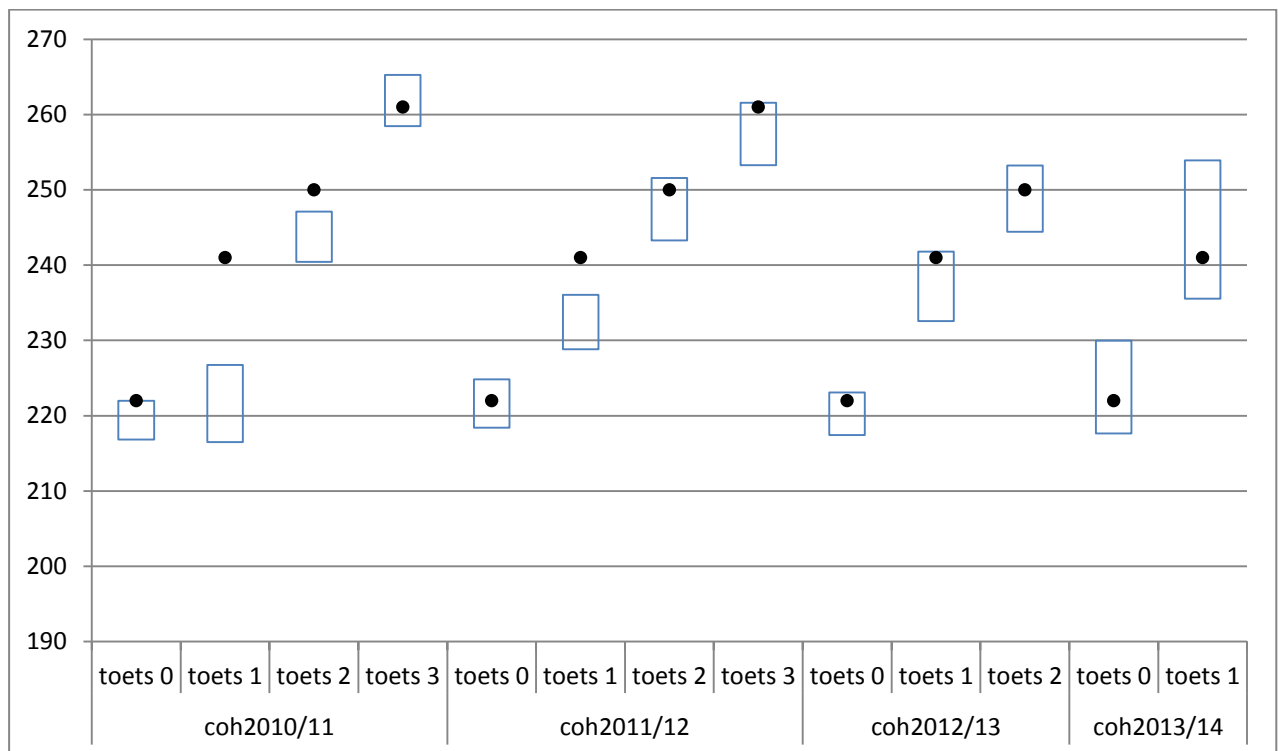
Een belangrijk verschil tussen de toetsen is dat de eerste drie de mogelijkheid bieden om de scores van leerlingen op opeenvolgende toetsen op dezelfde schaal uit te drukken, waarmee het mogelijk wordt om vooruitgang (leerwinst) in beeld te brengen. De laatste twee toetsen zijn gekoppeld aan de referentieniveaus voor taal en rekenen en geven een veel globaler beeld in termen van wel/niet beheersing van een bepaald referentieniveau.

Om na te gaan of leerlingen met verschillende kenmerken en achtergronden ook een lager of hoger beginniveau hebben, werd in de schoolrapporten eerst een beschrijving gegeven van de gemiddelde scores op de in schooljaar 2012/13 afgenomen begintoets op de school, onderscheiden naar de verschillende leerlingkenmerken. Daarmee kunnen de scholen zicht krijgen op de vraag of er bij de beoordeling van hun onderwijsopbrengsten rekening zou moeten worden gehouden met bepaalde kenmerken van hun leerlingenpubliek.

Als volgende stap in de beschrijvende analyses kregen scholen een overzicht van de gemiddelde toetsresultaten van verschillende cohorten leerlingen in hun school. Voor leerlingen, die aan opeenvolgende toetsen hebben deelgenomen, werd ook de gemiddelde leerwinst tussen die twee toetsafnames vermeld. Ter vergelijking waren ook de landelijke normen van de betreffende toets in het overzicht opgenomen. Daaruit kan afgeleid worden hoe leerlingen zich (landelijk gemiddeld) ontwikkelen tussen twee toetsafnames en dus wat de gemiddelde leerwinst is. De ontwikkeling van de gemiddelde resultaten op de toetsen in verschillende cohorten werd ook grafisch weergegeven. Daarbij werd tevens in beeld gebracht welke onzekerheidsmarges er rondom deze gemiddelden bestaan. De toetsen zijn immers geen perfecte meetinstrumenten: als dezelfde groep leerlingen de toets enkele dagen eerder of later zou maken dan zal daar nooit precies hetzelfde gemiddelde uitkomen. Deze onzekerheid is groter naarmate het aantal leerlingen kleiner is en naarmate er meer variatie in scores is tussen de leerlingen.

Hierna wordt een voorbeeld van zulke grafieken getoond. Het schoolgemiddelde met onzekerheidsmarges wordt daarin weergegeven als een blokje en de landelijke (norm)gemiddelden als een punt. In de voorbeeldgrafiek worden scores Nederlandse leesvaardigheid op de Cito-VVO toets weergegeven in vier opeenvolgende cohorten.

Voorbeeldfiguur 3.1 Nederlandse leesvaardigheid (Cito VVO) op 4 toetsmomenten bij 4 cohorten
(de stip geeft de landelijke norm weer; het blokje geeft het schoolgemiddelde met een onzekerheidsmarge naar boven en beneden)



In de grafiek staan de schoolscores van vier cohorten leerlingen. De cohorten 2010/11 en 2011/12 hebben alle beschikbare toetsen afgenomen. De andere twee cohorten nog niet omdat de leerlingen tijdens de laatste meting van de pilot pas in het tweede (cohort 2012/13) of eerste leerjaar (cohort 2013/14) zaten. We zien in de grafiek een aantal opvallende zaken:

- Leerlingen uit het cohort 2010/11 hebben aan het begin van het eerste leerjaar (toets 0) gemiddeld een score die nog net gelijk is aan het landelijk gemiddelde. Aan het einde van het eerste leerjaar (toets 1) scoren de leerlingen van de voorbeeldschool gemiddeld lager dan landelijk. In het tweede leerjaar (toets 2) lopen ze weer deels in en in het derde leerjaar (toets 3) ligt de score op het niveau van het landelijk gemiddelde. Voorgaande betekent dat de leerlingen op deze school in het eerste leerjaar gemiddeld minder leerwinst boeken dan landelijk, terwijl hun leerwinst in het tweede en derde leerjaar groter is dan landelijk.
- Ook leerlingen uit het cohort 2011/12 boeken in het eerste leerjaar gemiddeld minder leerwinst dan landelijk. Hier ligt de score in het tweede leerjaar al op het landelijk gemiddelde.
- Leerlingen uit de cohorten 2012/13 en 2013/14 scoren op alle toetsen die zij hebben gemaakt gemiddeld gelijk aan het landelijk gemiddelde en boeken ongeveer evenveel leerwinst als landelijk.
- Een interessante vraag voor deze school is of de school de afgelopen jaren veranderingen in de onderbouw in het onderwijs in Nederlandse leesvaardigheid heeft doorgevoerd. Deze zouden wellicht samen kunnen hangen met het niet meer optreden van een kleinere leerwinst dan landelijk bij de meest recente cohorten.

De toelichting in het kader onder de grafiek laat zien dat de school uit zo'n grafiek kan aflezen hoe de eigen leerlingen scoren ten opzichte van de landelijke gemiddelden, of zij in dezelfde mate vooruitgaan als landelijk én of er (opvallende) verschillen optreden tussen de opeenvolgende instroomcohorten.

Eindexamencijfers

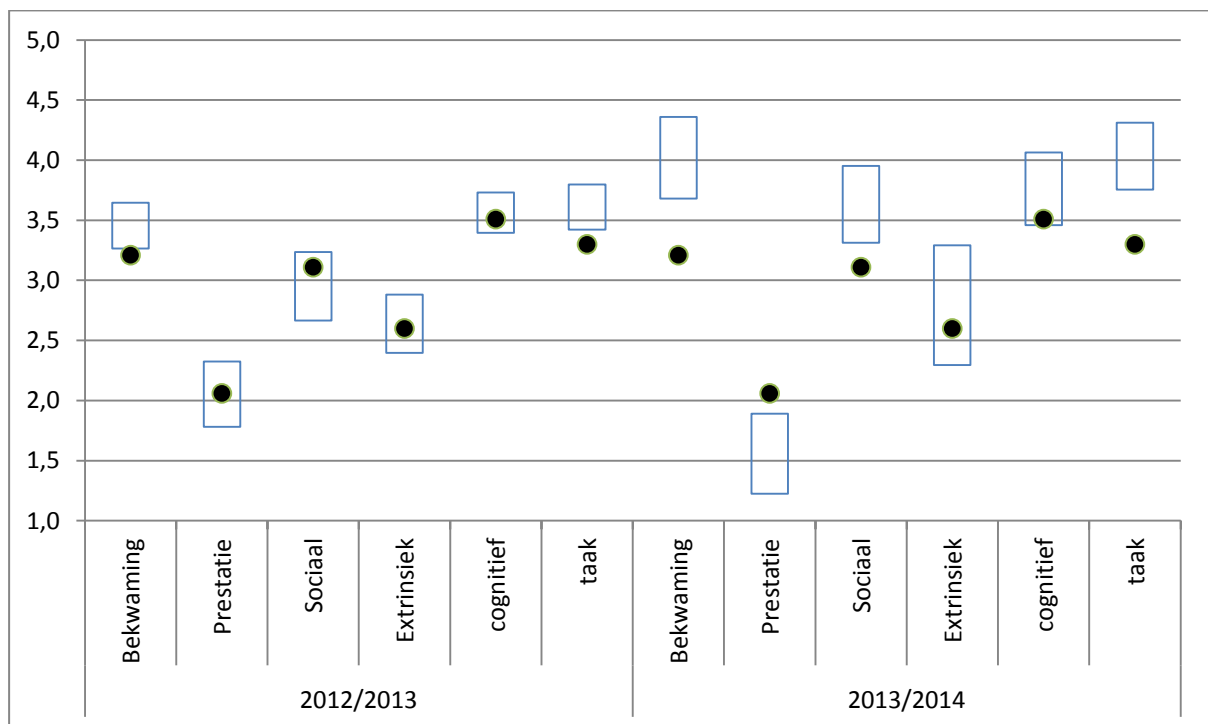
Bij de eindexamencijfers zijn twee soorten beschrijvende analyses naar de scholen gerapporteerd. De eindexamencijfers werden in verband gebracht met de (achtergrond)kenmerken van de leerlingen. Dat is van belang voor de vraag of zulke verschillen in leerlingkenmerken een rol zouden moeten spelen bij het beoordelen van de opbrengsten van de school. Verder werd, voor de leerlingen waarbij ook een score op een begintoets bekend is, een relatie gelegd tussen hun aanvangsscore en de eindexamenresultaten. Daarmee kan duidelijk worden of de bij aanvang sterkere of zwakkere leerlingen ook aan het eind van de schoolperiode betere of minder goede resultaten behalen. En dat is van belang voor de vraag of er bij de beoordeling van de eindopbrengsten van scholen rekening zou moeten worden gehouden met de (cognitieve) verschillen tussen leerlingen bij instroom.

We merken wel op dat deze beschrijvende analyses alleen konden worden uitgevoerd voor leerlingen die ook in de onderzochte schooljaren daadwerkelijk aan het eindexamen hebben deelgenomen. Het is goed mogelijk dat tussentijds uitgevallen leerlingen, bijvoorbeeld door afstroom, sterk verschillen van de leerlingen die wel tot aan het eindexamen zijn doorgestroomd in de school.

Vragenlijst Motivatie

Deze vragenlijst wordt gebruikt in het landelijke cohortonderzoek COOL⁵⁻¹⁸. Uit dat onderzoek zijn ook de landelijke gemiddelden (uit groep 8 basisonderwijs en leerjaar 3 voortgezet onderwijs) afkomstig op de vijf schalen, die in paragraaf 3.3 zijn omschreven. In de rapportage naar de scholen werden gemiddelde scores op deze schalen gegeven, in vergelijking met deze landelijk gemiddelden. Wanneer scholen twee opeenvolgende schooljaren de vragenlijst hebben afgenomen werden cijfers over beide jaren weergegeven. In de volgende voorbeeldfiguur worden gemiddelde schoolscores (opnieuw met onzekerheidsmarges) op de vijf schalen in twee schooljaren vergeleken met landelijke scores. Voor alle schalen lopen de scores van 1 tot 5.

Voorbeeldfiguur 3.2 Score op zes motivatieschalen (leerjaar 3 vmbootl/gl) in twee schooljaren; vergeleken met landelijke gemiddelden uit leerjaar 3 vmbo-tl/gl (de stip geeft de landelijke norm weer; het blokje geeft het schoolgemiddelde met een onzekerheidsmarge naar boven en beneden)



In de grafiek staan de scores van twee cohorten leerlingen in het derde leerjaar vmbo-tl/gl. Elk cohort leerlingen heeft de vragenlijst een keer ingevuld. Op basis van de grafiek kunnen we dus niets zeggen over de ontwikkeling ('leerwinst') van een cohort leerlingen. Wel kunnen we een vergelijking maken tussen de twee cohorten van de school en tussen elk cohort en de landelijke gemiddelden.

- De leerlingen uit het cohort 2012/13 scoren op de subschalen voor bekwaming en taakmotivatie gemiddeld boven het landelijk gemiddelde. Op de andere subschalen is de score van deze leerlingen gemiddeld gelijk aan het landelijk gemiddelde.
- De scores van de leerlingen uit het cohort 2013/14 laten een grilliger beeld zien. Deze leerlingen scoren op de subschalen voor bekwaming, sociale motivatie en taakmotivatie gemiddeld beter dan landelijk. Hun prestatie-motivatie is gemiddeld echter lager dan het landelijk gemiddelde. Op de overige twee subschalen scoren de leerlingen op het niveau van het landelijk gemiddelde.

In de schoolrapporten werden, ter vergelijking met de eigen school, gemiddelde scores uit het COOL-cohortonderzoek gepresenteerd (de stippen in voorbeeldfiguur 3.2). Deze scores komen steeds alleen uit leerjaar 3 van de onderscheiden niveaus van voortgezet onderwijs. Er is geen goed landelijk onderzoek beschikbaar, waarin ook de ontwikkeling op deze schalen van leerjaar tot leerjaar is gevolgd. Daarmee beschikken we dus niet over gegevens om na te gaan in hoeverre er in het algemeen (landelijk) sprake is van daling of stijging op deze schalen bij een herhaalde afname in het voortgezet onderwijs. Wel is uit (internationaal) onderzoek bekend dat er bij verschillende aspecten van motivatie sprake is van een daling, naarmate leerlingen langer hetzelfde soort onderwijs volgen (zie bijvoorbeeld Van der Veen & Peetsma, 2009).

De vragenlijsten zijn in beide onderzoeksjaren door een flink aantal leerlingen ingevuld (respectievelijk 547 en 713), maar er is een veel beperktere overlap van 187 leerlingen die ook in beide jaren hebben deelgenomen aan dit onderdeel van de pilot. Bij de scholen was kennelijk meer belangstelling voor een 'dwarsdoorsnede' (bijvoorbeeld vragen als: hoe scoren in beide schooljaren mijn eerstklassers op motivatie?) dan voor het nagaan van de ontwikkeling van de motivatie van hun leerlingen van jaar tot jaar (longitudinaal). Voor de leerlingen, die wel aan twee opeenvolgende metingen hebben meegedaan, kon ook iets gezegd worden over veranderingen in de scores op de schalen. In de schoolrapporten is het verschil tussen de scores bij beide afnamen berekend en ingedeeld als:

- gedaald: 0.2 punt of meer omlaag;
- gelijk gebleven: minder dan 0.2 omhoog of omlaag;
- gestegen: 0.2 punt of meer omhoog.

Vervolgens werden per schaal de aantallen gegeven van deze drie groepen. De volgende tabel geeft een voorbeeld voor één van de pilotscholen.

Voorbeeldtabel 3.1 Veranderingen in scores motivatie; (aantallen)

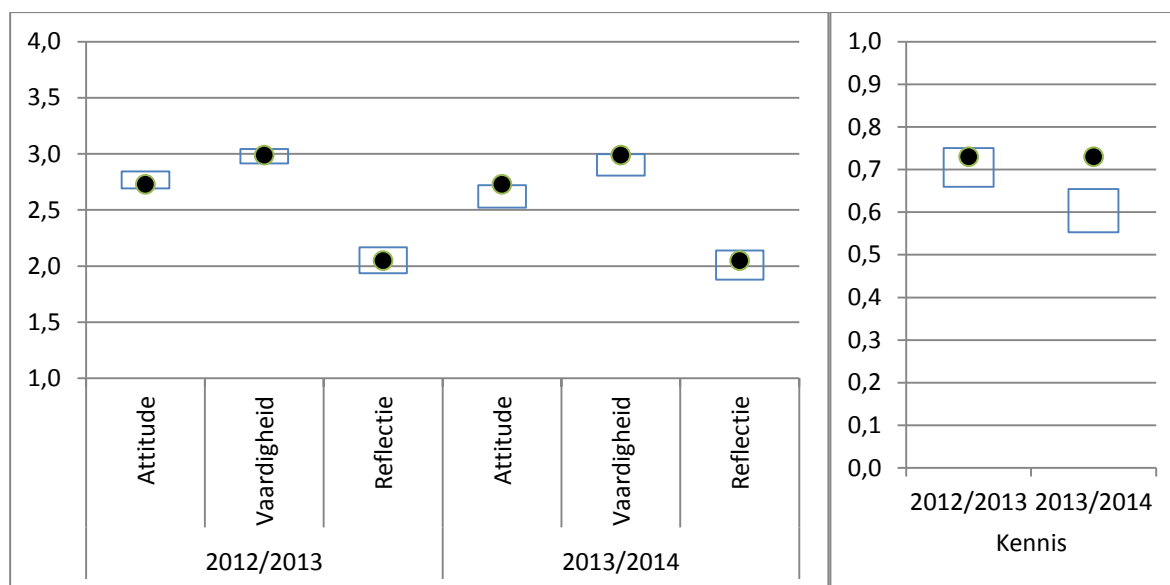
	Bekwaming (Mastery)	Prestatie (Perform.)	Sociaal (Social)	Extrinsiek (Extrinsic)	Cognitief zelfvertr.	Taak- motivatie
gedaald	2	11	5	11	6	5
gelijk	8	5	3	2	6	4
gestegen	7	1	9	4	6	9

Vragenlijst Burgerschapscompetenties

Ook deze vragenlijst wordt gebruikt in het landelijke cohortonderzoek COOL⁵⁻¹⁸. Het instrument levert scores op vier schalen, die in paragraaf 3.3 zijn omschreven. In de rapportage naar de scholen werden gemiddelde scores op deze schalen gegeven, in vergelijking met landelijk gemiddelden uit leerjaar 3 van het voortgezet

onderwijs. Wanneer scholen twee opeenvolgende schooljaren de vragenlijst hebben afgenomen werden cijfers over beide jaren weergegeven. In de volgende voorbeeldfiguur worden gemiddelde schoolscores (met onzekerheidsmarges) op de vijf schalen in twee schooljaren vergeleken met landelijke scores. Voor drie schalen lopen de scores van 1 tot 4. Bij de kennis-schaal loopt de score van 0 tot 1 (de proportie juiste antwoorden).

Voorbeeldfiguur 3.3 Score op vier burgerschapsschalen (leerjaar 3 kader) in twee schooljaren; vergeleken met landelijke gemiddelden uit leerjaar 3 kader (de stip geeft de landelijke norm weer; het blokje geeft het schoolgemiddelde met een onzekerheidsmarge naar boven en beneden)



In de grafiek staan de scores van twee cohorten leerlingen in het derde leerjaar vmbo-kader. Elk cohort leerlingen heeft de vragenlijst een keer ingevuld. Op basis van de grafiek kunnen we dus niets zeggen over de ontwikkeling ('leerwinst') van een cohort leerlingen. Wel kunnen we een vergelijking maken tussen de twee cohorten van de school en tussen elk cohort en de landelijke gemiddelden.

- De leerlingen uit beide cohorten scoren op de competenties 'attitude', 'vaardigheid' en 'reflectie' gemiddeld op het niveau van het landelijk gemiddelde.
- Voor de leerlingen uit cohort 2012/13 geldt dat ook voor kennis. De kennis over burgerschap van de leerlingen uit het cohort 2013/14 blijft achter bij het landelijk gemiddelde. Dit laatste kan een aandachtspunt zijn voor het onderwijs aan dit cohort leerlingen.

Ook hier zijn de vragenlijsten maar door een beperkt aantal leerlingen in beide onderzoeksjaren ingevuld (127). Alleen voor die leerlingen kan dan ook iets gezegd worden over veranderingen in de scores op de schalen. In de schoolrapporten is het verschil tussen de scores bij beide afnamen berekend en ingedeeld als:

- gedaald: 0.2 punt of meer omlaag (bij de schaal Kennis: 0.1 punt);
- gelijk gebleven: minder dan 0.2 omhoog of omlaag (bij de schaal Kennis: 0.1 punt);
- gestegen: 0.2 punt of meer omhoog (bij de schaal Kennis: 0.1 punt).

Vervolgens werden per schaal de aantallen gegeven van deze drie groepen. De volgende tabel geeft voor één deelnemende school een illustratie daarvan.

Voorbeeldtabel 3.2 Veranderingen in scores burgerschap; (aantallen)

	Attitude	Vaardigheid	Reflectie	Kennis
leerjaar 1 → 2				
gedaald	26	26	28	22
gelijk	32	29	22	26
gestegen	10	13	15	12
leerjaar 2 → 3				
gedaald	14	11	17	5
gelijk	15	21	14	18
gestegen	15	11	11	16

En hoewel er dus, ter vergelijking, wel landelijke gegevens uit leerjaar 3 bekend zijn, ontbreekt onderzoek dat de ontwikkeling van leerlingen van jaar tot jaar in beeld brengt. We kunnen dus ook hier niet goed nagaan in hoeverre er in het algemeen (landelijk) sprake is van daling of stijging op deze schalen bij een herhaalde afname in het voorgezet onderwijs.

3.5 Slotbeschouwing

Zoals eerder opgemerkt, verschilden de uitgevoerde beschrijvende analyses van school tot school, afhankelijk van de beschikbare gegevens voor de betreffende school. Voor elke school was er dan ook in beide onderzoeksjaren een uniek schoolrapport, dat uitgebreid met de school werd besproken.

Uit de per school verschillende analyses valt echter wel een grote lijn te destilleren, namelijk dat het bij de meeste van de gebruikte toetsscores mogelijk was om de gemiddelde leerwinst van (groepen) leerlingen goed in beeld te krijgen. Dat geldt in elk geval voor de toetsen, waarvan de scores zijn uit te drukken op een doorlopende vaardigheidsschaal. Voor toetsen, die alleen de beheersing van de referentieniveaus voor taal en rekenen weerspiegelen, is leerwinst veel minder goed aan te geven. De mate van vooruitgang van leerlingen tussen, bijvoorbeeld, niveau 2F en 3F valt niet nauwkeurig te bepalen. Ook voor eindexamencijfers geldt dat er niet echt leerwinst kan worden bepaald, maar dat alleen kan worden nagegaan of bij aanvang zwak scorende leerlingen toch goede eindresultaten behaalden of niet. Met daarbij de restrictie dat het dan alleen gaat om de leerlingen die steeds op dezelfde school zijn gebleven.

Voor de sociaal-emotionele opbrengsten van onderwijs (motivatie en burgerschapscompetenties) geldt dat deze wel in beeld konden worden gebracht, maar dat de mate van ontwikkeling van de leerlingen op deze domeinen niet goed kan worden afgezet tegen landelijke gegevens. De meeste scholen hebben deze instrumenten dan ook gebruikt om een beeld te krijgen van een bepaalde jaargroep en niet om de ontwikkeling van leerlingen te volgen.

4 Prestatieniveau en leerwinst vergeleken

4.1 Inleiding

Zoals in paragraaf 3.1 aangegeven kunnen de opbrengsten van scholen zowel beoordeeld worden naar het gemiddelde prestatieniveau, dat de leerlingen op een bepaald moment halen, als naar de leerwinst die de leerlingen over de afgelopen periode hebben behaald. Het is aan te bevelen om beide aspecten te gebruiken bij het beoordelen van onderwijsopbrengsten (Roeleveld e.a., 2008; Bosker, 2012). Op die manier kunnen scholen in een van de volgende kwadranten ingedeeld worden:

laag niveau veel leerwinst	hoog niveau veel leerwinst
laag niveau weinig leerwinst	hoog niveau weinig leerwinst

Met de gegevens uit de pilot wordt het mogelijk om de deelnemende scholen op een dergelijke manier in te delen. Wij doen dit in paragraaf 4.2 per type afdeling en met, als voorbeeld, de score op Nederlandse leesvaardigheid uit de Cito-VVO toetsen voor het eerste leerjaar. Op de horizontale as van de volgende figuren staat de gemiddelde score aan het eind van het eerste leerjaar. Op de verticale as staat de leerwinst, behaald tussen het begin en het eind van dat eerste leerjaar. De gekruiste lijnen door het midden geven de landelijke gemiddelden uit het normeringsonderzoek van Cito. De afzonderlijke scholen worden in deze figuren anoniem aangeduid met een letter.

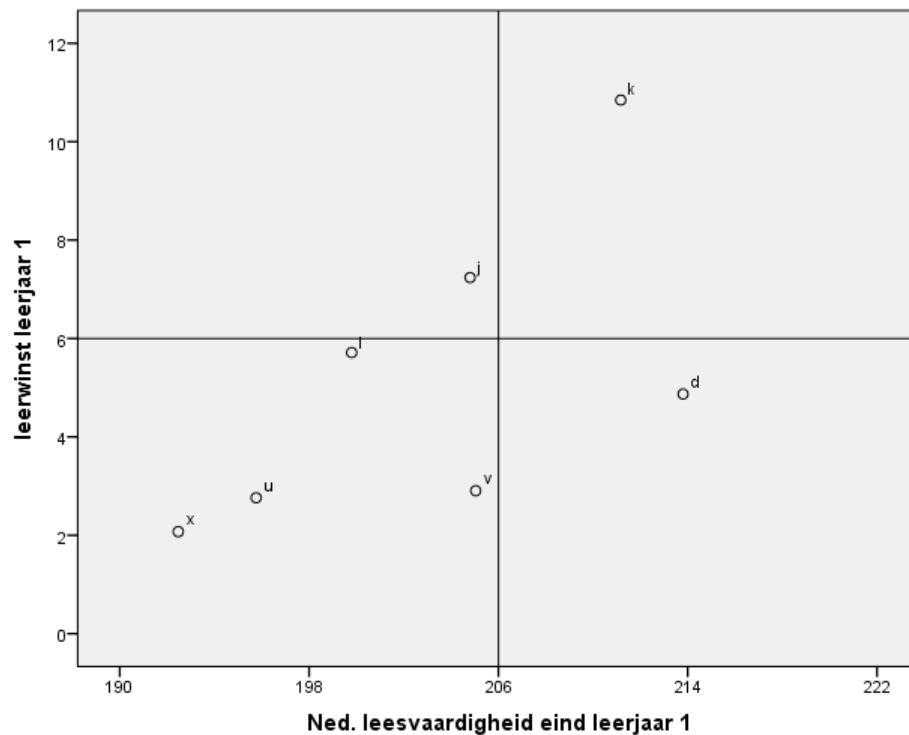
Voor de figuren gebruiken we zoveel mogelijk gegevens over de vooruitgang van leerlingen in het eerste leerjaar, niet alleen uit de onderzoeksjaren van de pilot maar ook historische gegevens van eerdere schooljaren (voor zover bekend). De schoolgemiddelden worden als een punt (met letteraanduiding) weergegeven. Rondom deze gemiddelden zit echter ook een (statistische) onzekerheidsmarge. Onder de figuren vermelden we of scholen significant van de landelijke gemiddelde normen afwijken; als dat niet het geval is wordt de afkorting n.s. vermeld. Daarbij blijkt dat het vaak niet heel scherp mogelijk is om de scholen in de vier kwadranten in te delen: de onzekerheid in de metingen verhindert dit dan.

4.2 Prestatieniveau en leerwinst in de pilot

Vmbo-basis

In de volgende figuur laten we de gemiddelde toetsscore aan het eind van leerjaar 1 zien op de horizontale as, terwijl de gemiddeld bereikte leerwinst in dat eerste leerjaar op de verticale as is uitgezet.

Figuur 4.1 *Score leesvaardigheid eind leerjaar 1 versus leerwinst in leerjaar 1; voor vmbo-basis (normering Toets 0 en Toets 1: 200 resp. 206; vooruitgang 6)*

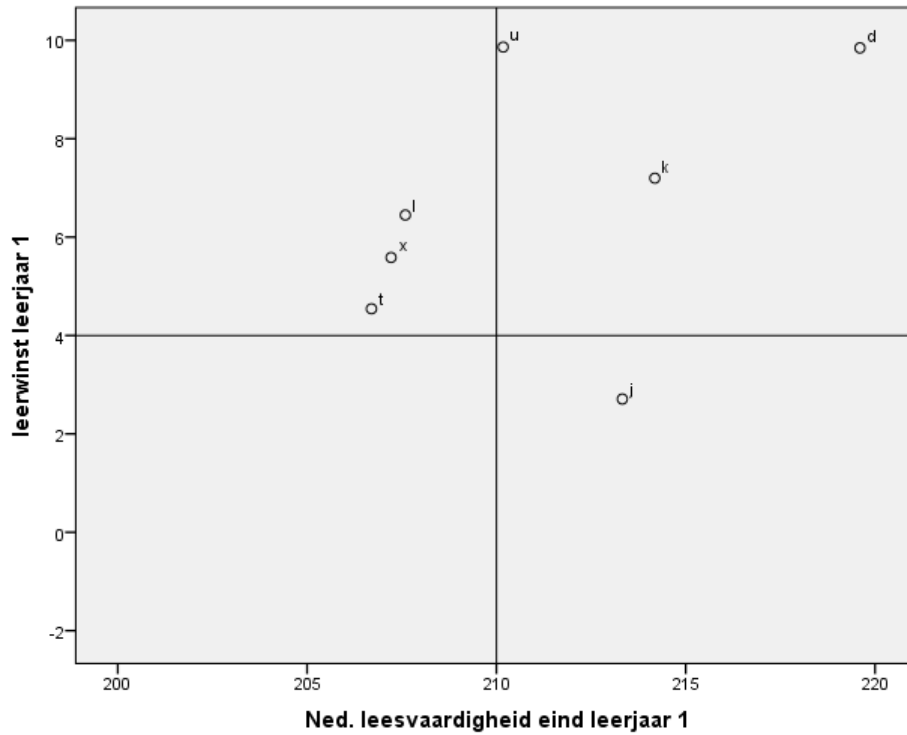


school	d	j	k	l	u	v	x
eind leerjaar 1	hoger	n.s.	hoger	lager	lager	n.s.	lager
leerwinst	n.s.	n.s.	hoger	n.s.	n.s.	lager	n.s.

School k valt in het meest gunstige kwadrant: de prestaties aan het eind van leerjaar 1 zijn bovengemiddeld, evenals de leerwinst die in dat jaar behaald is. Voor scholen x en u geldt juist het tegendeel: de twee indicatoren zijn ondergemiddeld. Voor beide scholen geldt echter ook dat de afwijking van leerwinst van het landelijke gemiddelde niet significant is.

Vmbo-kader

Figuur 4.2 Score leesvaardigheid eind leerjaar 1 versus leerwinst in leerjaar 1; voor vmbo-kader (normering Toets 0 en Toets 1: 206 resp. 210; vooruitgang 4)

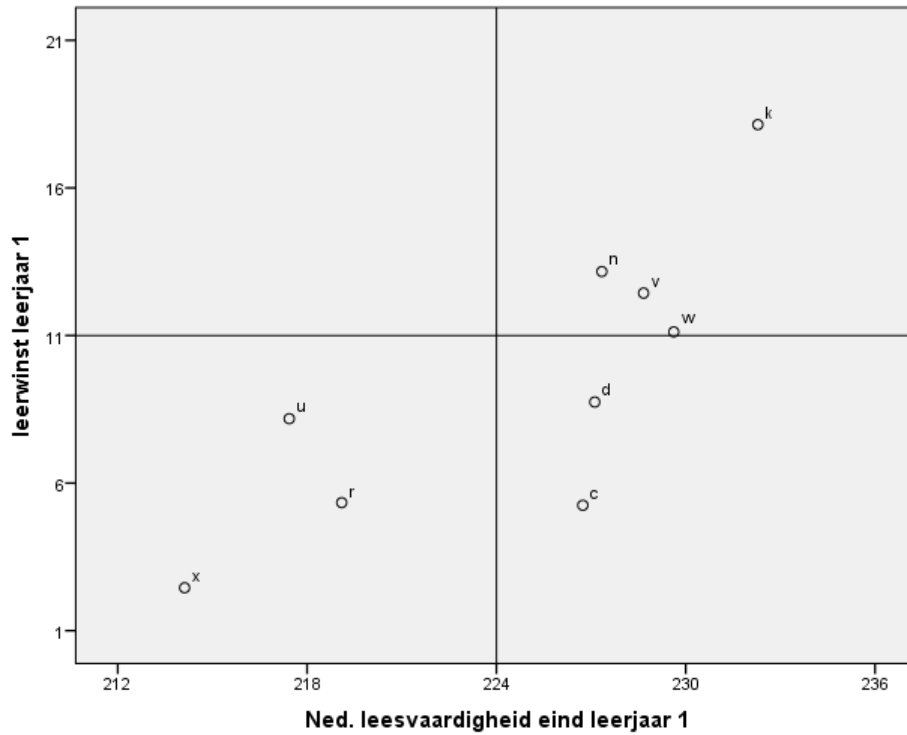


school	d	j	k	l	t	u	x
eind leerjaar 1	hoger	n.s.	hoger	n.s.	lager	n.s.	lager
leerwinst	hoger	n.s.	hoger	n.s.	n.s.	hoger	hoger

Hier valt vooral school d op: zowel de eindscore als de leerwinst liggen flink boven het gemiddelde. Op deze zelfde school behaalden de vmbo-basis leerlingen ook een bovengemiddelde score, maar was de leerwinst ongeveer gemiddeld (zie Figuur 4.1). Er valt hier geen enkele school in het meest ongunstige kwadrant linksonder.

Vmbo-tl/gl

Figuur 4.3 Score leesvaardigheid eind leerjaar 1 versus leerwinst in leerjaar 1; voor vmbo-tl/gl (normering Toets 0 en Toets 1: 213 resp. 224; vooruitgang 11)

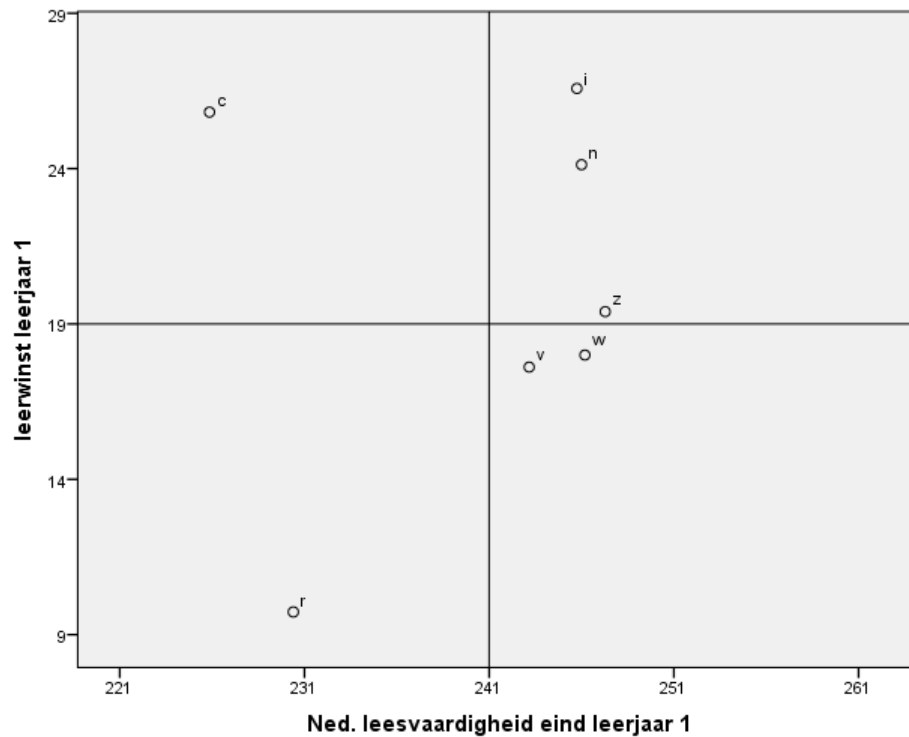


school	c	d	k	n	r	u	v	w	x
eind leerjaar 1	n.s	n.s.	hoger	hoger	lager	lager	hoger	hoger	lager
leerwinst	n.s.	n.s.	hoger	n.s.	lager	n.s.	n.s.	n.s.	lager

De meeste scholen vallen in de kwadranten rechtsboven (zowel eindscore als leerwinst gunstig) of linksonder (beide ongunstig), hoewel de verschillen met de landelijke normen niet altijd significant blijken te zijn. Scholen d en c zitten in het kwadrant rechtsonder, maar hun scores wijken niet significant af van de landelijke normen.

Havo

Figuur 4.4 Score leesvaardigheid eind leerjaar 1 versus leerwinst in leerjaar 1; voor havo (normering Toets 0 en Toets 1: 222 resp. 241; vooruitgang 19)

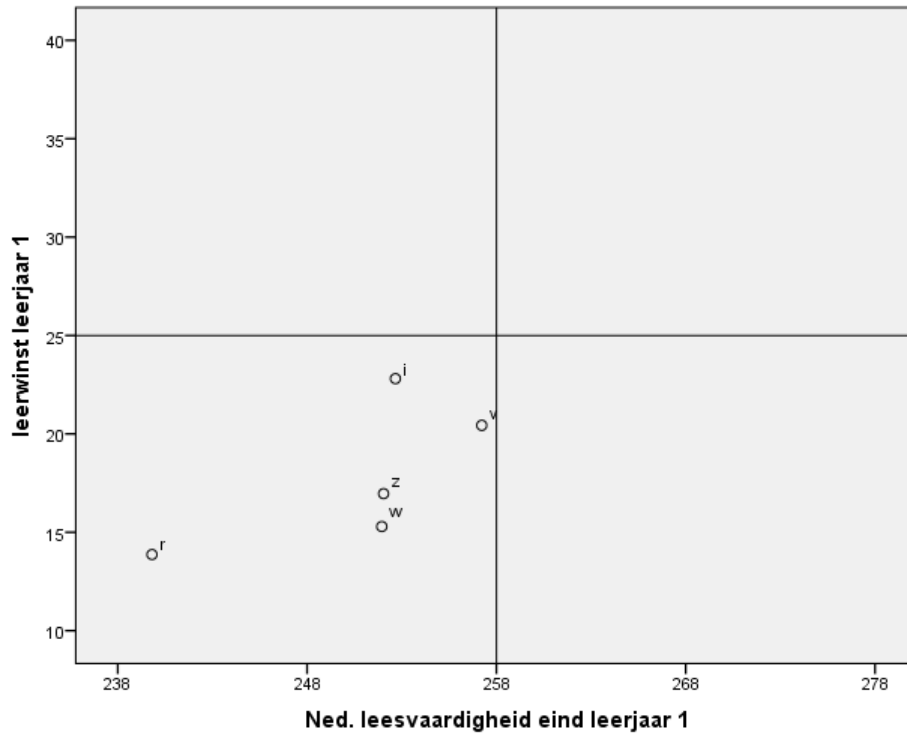


school	c	i	n	r	v	w	z
eind leerjaar 1	lager	hoger	hoger	lager	n.s.	hoger	hoger
leerwinst	hoger	hoger	hoger	lager	n.s.	n.s.	n.s.

Hier valt vooral school c in het kwadrant linksboven op. Bij deze school is het niveau aan het eind van leerjaar 1 lager dan landelijk gemiddeld, maar tegelijk wordt er wel meer leerwinst behaald dan landelijk. Kennelijk zijn hier leerlingen met lage scores aan het begin van leerjaar 1 ingestroomd, die weliswaar relatief veel leerwinst hebben behaald, maar aan het eind van leerjaar 1 nog steeds achterblijven bij het landelijk gemiddelde.

Vwo

Figuur 4.5 Score leesvaardigheid eind leerjaar 1 versus leerwinst in leerjaar 1; voor vwo (normering Toets 0 en Toets 1: 233 resp. 258; vooruitgang 25)



school	i	r	v	w	z
eind leerjaar 1	lager	lager	n.s.	lager	lager
leerwinst	n.s.	lager	n.s.	lager	lager

Opvallend is dat alle vijf scholen in het kwadrant linksonder liggen: laag niveau en weinig leerwinst. Overigens zijn de verschillen tussen schoolgemiddelde en landelijk cijfers niet steeds significant. Maar voor scholen r, w en z geldt dat zij op beide indicatoren voor schoolopbrengst significant lager scores dan landelijk gemiddeld.

4.3 Slotbeschouwing

De analyses in dit hoofdstuk laten zien dat wanneer we scholen beoordelen naar de leerwinst die de leerlingen in het eerste schooljaar boeken, dat een heel ander beeld kan geven dan wanneer we alleen kijken naar het bereikt prestatieniveau van de leerlingen aan het eind van dat schooljaar. Beide aspecten van de opbrengsten van onderwijs zouden dan ook bij de beoordeling van scholen in ogenschouw moeten worden genomen (Bosker, 2012).

Tot slot merken we op dat de indeling van scholen in deze vier kwadranten laat zien waar de scholen staan, in vergelijking met landelijke normen voor de gebruikte toetsen. Maar de verschillen, die we tussen de scholen zien, kunnen niet rechtstreeks en eenvoudig worden toegeschreven aan verschillen in kwaliteit van de scholen. Ze kunnen ook te maken hebben met een hoog of laag instroomniveau van de leerlingen en/of met allerlei andere kenmerken van leerlingen en de context van de school, die het leren en vooruitgaan belemmeren of

juist bevorderen. Daarmee komen we bij het begrip toegevoegde waarde: welk deel van de vooruitgang van leerlingen kan beschouwd worden als unieke bijdrage van de school zelf en niet aan de meer of minder gunstige kenmerken van de leerlingen of de context van de school?

5 Toegevoegde waarde

5.1 Inleiding

Bij het bepalen van toegevoegde waarde wordt rekening gehouden met verschillen tussen scholen in hun leerlingenpubliek. Bij sommige leerlingen is het makkelijker om vooruitgang te boeken dan bij andere leerlingen. Bij leerlingen met ‘ongunstige’ achtergrondkenmerken zullen de verwachtingen omtrent de te behalen leerwinst lager zijn dan bij andere leerlingen. En een school met veel leerlingen met ‘ongunstige’ achtergrondkenmerken zal meer moeite moeten doen om de landelijke normen, gebaseerd op ‘doorsnee’ leerlingen, te behalen dan een school met veel leerlingen met juist ‘gunstige’ achtergrondkenmerken. Voor de bepaling van welke verwachtingen we bij welke leerlingen mogen hebben, zijn geen rechtstreekse landelijke vergelijkingsgegevens voorhanden. Deze moeten gehaald worden uit onderzoek waarin de resultaten van scholen met verschillende leerlingpopulaties onderling worden vergeleken.

In het meeste onderzoek naar de toegevoegde waarde van scholen wordt uitgegaan van een dataverzameling met gestandaardiseerde meetinstrumenten op een representatieve steekproef van scholen. In de analyses worden de opbrengsten van scholen (in termen van bereikt eindniveau of van leerwinst; met meer of minder uitgebreide correcties voor verschillen in leerlingpopulatie en andere kenmerken van de scholen) onderling vergeleken: zijn er scholen die met hun leerlingen gemiddeld duidelijk betere of juist mindere resultaten boeken dan de andere scholen in de steekproef? De statistische modellen kunnen daarbij eenvoudiger of complexer zijn, maar in de uitkomsten gaat het steeds om een relatieve vergelijking: er zijn scholen die bovengemiddeld scoren en scholen die achterblijven ten opzichte van het gemiddelde van alle scholen in de steekproef (zie ook hoofdstuk 3.1 over indicatoren op basis van toegevoegde waarde). De representatieve steekproef levert dus de benchmark, waartegen de toegevoegde waarde van afzonderlijke scholen kan worden afgezet.

Voorbeelden van onderzoek naar toegevoegde waarde in Nederland zijn verschillende deelstudies in het proefschrift van Timmermans (2012) op data van het VOCL-cohort, onderzoek naar groeicurven van leerlingen uit het PRIMA-onderzoek (Guldmond & Bosker, 2006; Roeleveld, 2011), onderzoek op COOL-data naar correctiefactoren in het basisonderwijs (Roeleveld e.a., 2011; Mooij e.a., 2012) en recent onderzoek naar indicatoren voor leeropbrengsten in het voortgezet onderwijs op VOCL-data en gegevens uit de landelijke onderwijsnummerbestanden (Rekers-Mombarg, e.a., 2014).

Bij de pilots rondom leerwinst en toegevoegde in het voortgezet onderwijs is van een dergelijke representatieve steekproef geen sprake. Er doet een beperkt aantal scholen mee, met afdelingen van verschillend niveau (van de basisberoepsgerichte leerweg tot aan gymnasium) en met uiteenlopende, onderling niet vergelijkbare meetinstrumenten. Deze situatie levert twee problemen op voor analyses rond toegevoegde waarde. Allereerst is er, als gezegd, geen goede landelijke steekproef waarmee vergeleken kan worden. Als een school in de analyses weinig toegevoegde waarde zou laten zien, kan dat evengoed komen doordat de school vergeleken wordt met enkele zeer goede scholen, die toevallig aan de pilot deelnemen, en niet met een representatieve groep scholen.

Maar daarnaast is er, door het gebruik van verschillende meetinstrumenten en scholen met verschillende niveaus van voortgezet onderwijs, ook steeds maar een beperkt aantal scholen beschikbaar voor analyse. Uit de onderzoeksliteratuur blijkt dat de schattingen van toegevoegde waarde een behoorlijk onzekerheidsmarge kennen en dat het vaak alleen mogelijk is om scholen met de allerlaagste en allerhoogste uitkomsten van

elkaar te onderscheiden. Het gros van de scholen (soms 60 tot 80 procent) verschilt niet significant van elkaar in toegevoegde waarde (Raudenbush, 2004).¹⁵

Bij de kleine aantallen scholen met eenzelfde afdeling en eenzelfde toets in deze pilot verwachten we nauwelijks scholen significant van elkaar te kunnen onderscheiden. Dit betekent dat in het onderzoek een onderlinge vergelijking van afdelingen en scholen maar beperkt zinvol is. Om deze reden gaan we in dit hoofdstuk meer beschrijvend verder in op de grondgedachte achter het idee van de toegevoegde waarde van scholen: scholen verschillen in hun leerlingenpubliek en bij de beoordeling van hun opbrengsten moet rekening worden gehouden met het feit dat sommige leerlingen makkelijker en andere moeilijker vooruitgang zullen boeken. We doen dat niet door met statistische modellen te corrigeren voor verschillen in leerlingpubliek tussen de scholen, maar door meer beschrijvend te laten zien of en in hoeverre de in de pilot beschikbare leerlingkenmerken samenhangen met zowel het instroomniveau van de leerlingen, als met hun leerwinst. Waar in onderzoek naar toegevoegde waarde een reeks correctiefactoren (dat kunnen zowel leerling- als contextkenmerken zijn) in een complex statistisch model gecombineerd worden, knippen we hier als het ware de vraag naar toegevoegde waarde op in stukjes, per leerlingkenmerk.

In de volgende paragrafen gaan we in op het beginniveau en de leerwinst in het eerste leerjaar. Daarbij kijken we niet meer naar afzonderlijke scholen, maar gebruiken gegevens van alle leerlingen op de pilotscholen. Wel maken we een onderscheid naar het niveau van onderwijs, dat de leerlingen volgen. We kiezen voor gegevens van de Cito-VVO-toetsen, omdat daar de meeste gegevens over beschikbaar zijn. We laten steeds de beginscores zien van leerlingen in de verschillende onderwijstypen en de leerwinst die zij in dat eerste leerjaar boeken.¹⁶

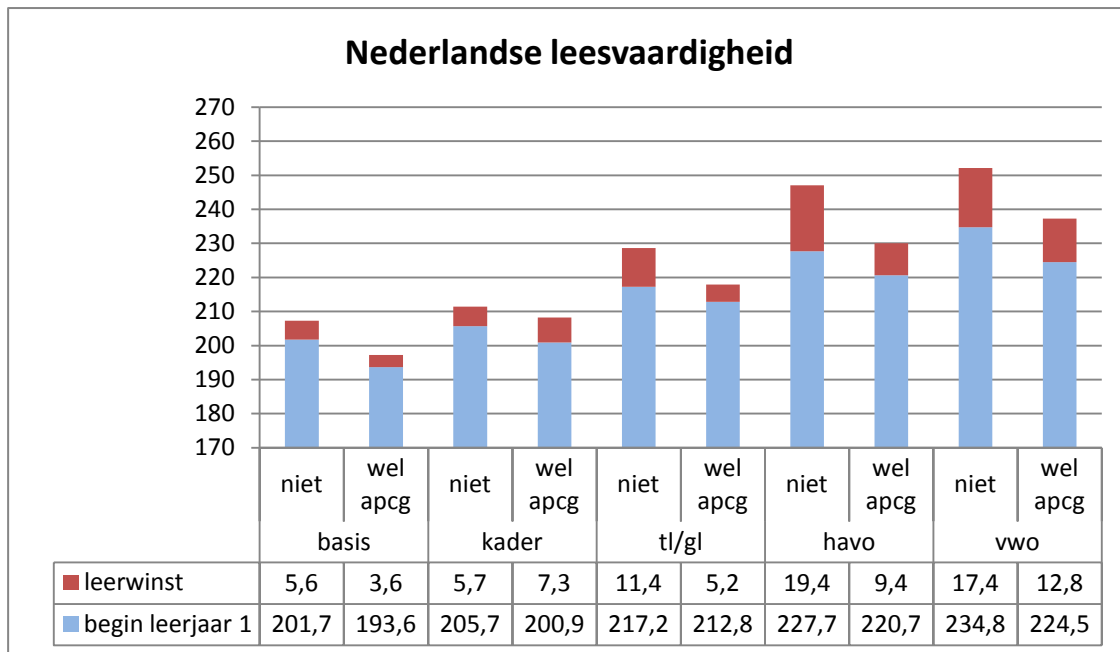
5.2 Beginniveau en leerwinst naar afkomst uit APCG-gebied

In het voortgezet onderwijs krijgen scholen, in het kader van het zogenaamde Leerplusarrangement extra middelen als er veel leerlingen afkomstig zijn uit een armoedeprobleemcumulatiegebied (APCG). De gedachte daarachter is dat er bij die leerlingen een grote kans is dat er sprake is van onderwijsachterstanden en een weinig stimulerend thuismilieu (Claassen e.a., 2005). In de navolgende figuren laten we zien of leerlingen uit een APCG-gebied met lagere toetscores in de school instromen én of hun leerwinsten anders zijn dan bij de overige leerlingen. We onderzoeken de vier scores uit de Cito-toetsen T0 en T1.

¹⁵ Zie ook verschillende figuren van schoolresiduals en betrouwbaarheidsintervallen bij Timmermans (2012).

¹⁶ Dergelijke analyses kunnen natuurlijk ook uitgevoerd worden voor de leerjaren twee en drie.

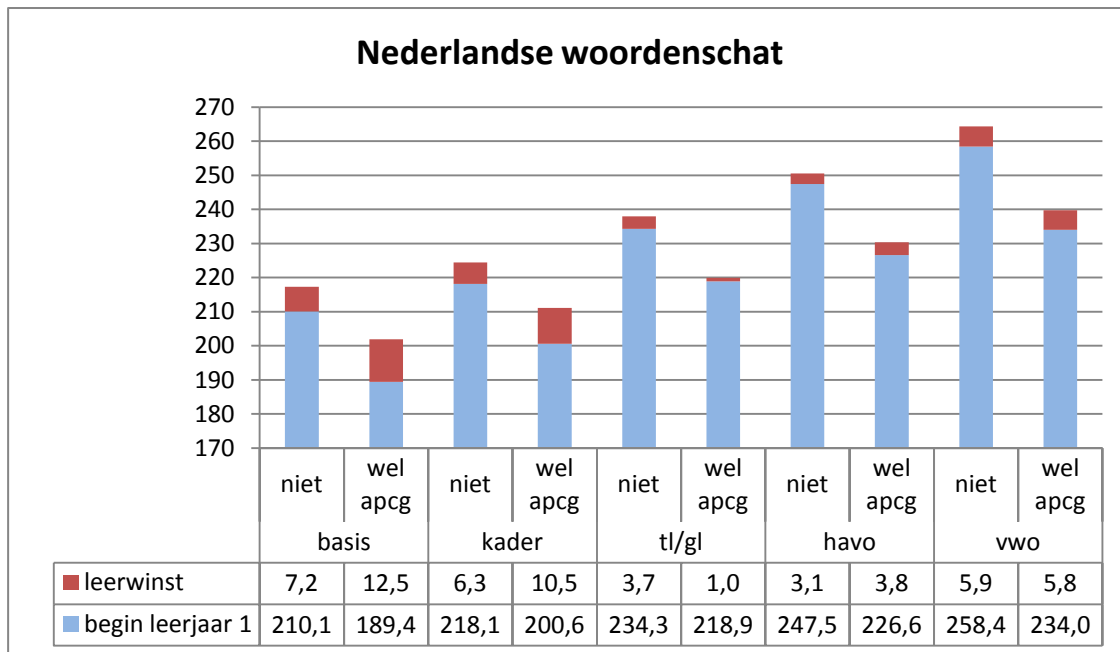
Figuur 5.1 Instroomniveau en leerwinst eerste leerjaar bij Nederlandse leesvaardigheid; naar afkomst uit APCG-gebied



Bij alle vijf schooltypen hebben leerlingen uit APCG-gebieden gemiddeld lagere scores bij binnenkomst dan de overige leerlingen. De verschillen zijn behoorlijk: effectgroottes¹⁷ van 0.4 tot 0.7. Verder blijkt dat ook de leerwinst doorgaans lager uitvalt. Alleen bij vmbo-kader behalen APCG-leerlingen wat meer leerwinst dan niet APCG-leerlingen. Maar over het algemeen zien we dus de aanvankelijke achterstand van APCG-leerlingen nog toenemen in het eerste leerjaar.

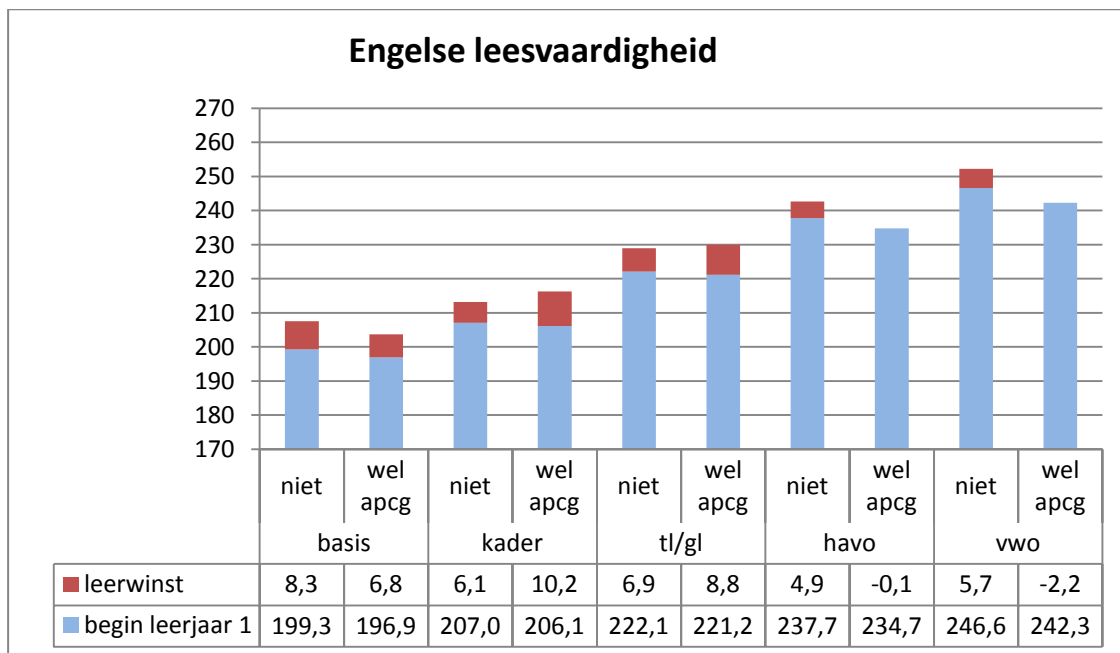
¹⁷ De effectgrootte geeft het verschil in gemiddelde gedeeld door de standaarddeviatie. Een verschil van 0.2 wordt in de literatuur beschouwd als klein, 0.5 als middelgroot en 0.8 en meer als groot (Cohen, 1988).

Figuur 5.2 *Instroomniveau en leerwinst eerste leerjaar bij Nederlandse woordenschat; naar afkomst uit APCG-gebied*



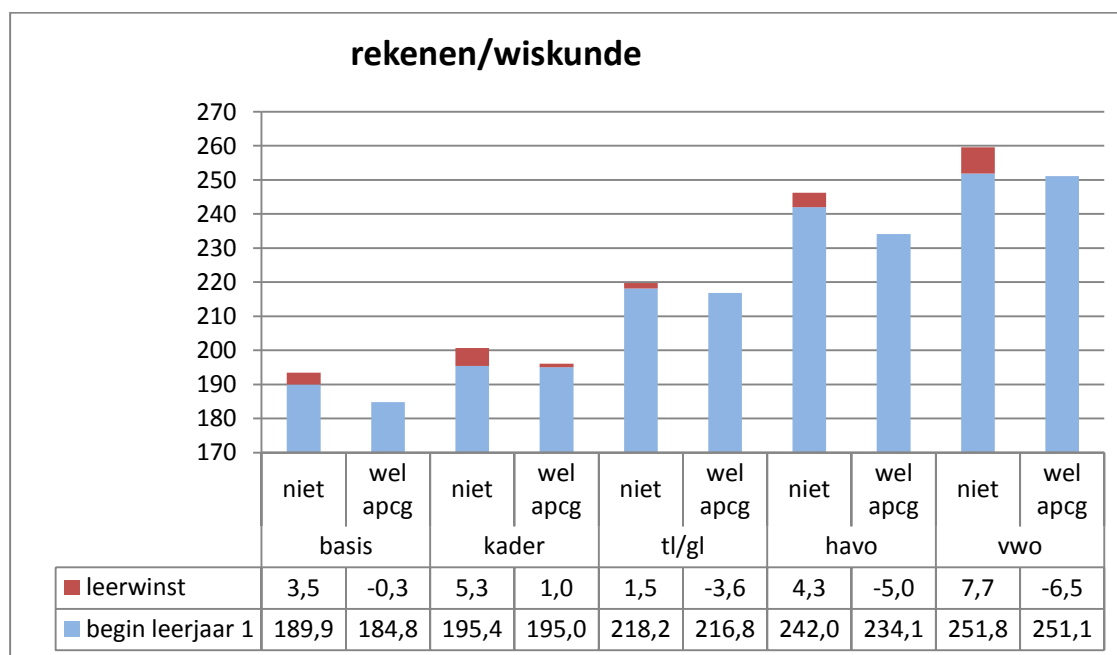
Ook bij Nederlandse woordenschat scoren APCG-leerlingen fors lager bij instroom dan de overige leerlingen (effectgrootte van 0.7 tot 1.0). De verschillen bij leerwinst zijn veel kleiner en zowel bij basis als bij kader boeken de APCG-leerlingen wat meer leerwinst in het eerste leerjaar dan niet-APCG-leerlingen. Zij lopen dus een deel van hun achterstand in, maar scoren ook aan het eind van het eerste leerjaar nog duidelijk lager.

Figuur 5.3 *Instroomniveau en leerwinst eerste leerjaar bij Engelse leesvaardigheid; naar afkomst uit APCG-gebied*



Bij Engelse leesvaardigheid scoren APCG-leerlingen wat lager, maar de verschillen zijn hier beperkt (het grootste verschil, bij vwo, heeft een effectgrootte van 0.16). Bij havo en vwo is er bij de APCG-leerlingen geen leerwinst, maar juist een lichte daling. De niet APCG-ers gaan juist wel vooruit: de verschillen in leerwinst met APCG-ers hebben een effectgrootte van 0.24 tot 0.33. Bij kader en tl/gl gaan de APCG-ers iets meer vooruit dan de overige leerlingen (effectgrootte rond de 0.15).

Figuur 5.4 Instroomniveau en leerwinst eerste leerjaar bij rekenen/wiskunde; naar afkomst uit APCG-gebied



Bij rekenen/wiskunde beginnen leerlingen uit APCG-gebieden doorgaans met een (lichte) achterstand. Alleen bij havo en basis zijn de verschillen vrij groot, met effectgroottes van resp. 0.5 en 0.3. Opvallend genoeg gaan de leerlingen uit APCG-gebieden (behalve bij kader) gedurende het eerste leerjaar achteruit in vaardigheid, terwijl de overige leerlingen wel leerwinst boeken. Bij de drie vmbo-niveaus ligt de effectgrootte van het verschil in leerwinst rond de 0.3, maar bij havo en vwo duidelijk hoger: 0.7 tot 0.8.

Het lijkt vreemd dat leerlingen gedurende het schooljaar achteruitgaan in een bepaalde vaardigheid. Dit heeft vermoedelijk te maken met het feit dat deze toets echt de vaardigheid rekenen meet, een vak dat doorgaans niet expliciet wordt onderwezen in het eerste leerjaar. Wanneer het rekenen niet echt systematisch wordt bijgehouden, is het goed mogelijk dat de betreffende vaardigheden na bepaalde tijd wegzakken.

Samenvattend

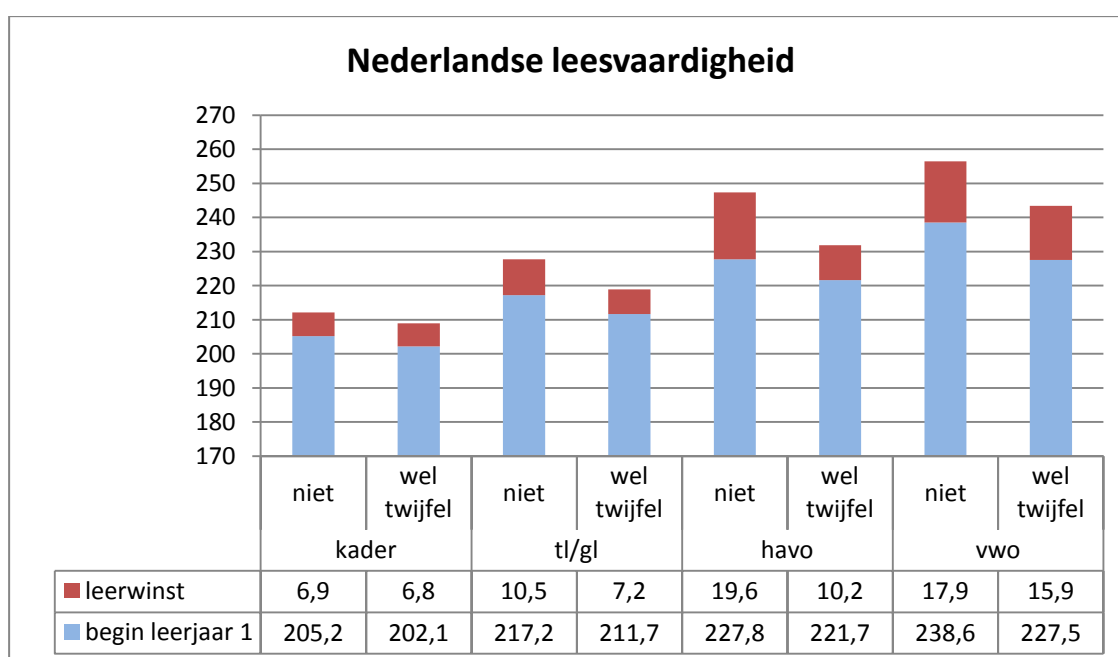
Leerlingen uit een APCG-gebied behalen bij instroom lagere scores; de verschillen met overige leerlingen zijn het kleinst bij Engelse leesvaardigheid. Bij leerwinst is het beeld meer divers. Bij sommige schooltypen en bij sommige vaardigheden halen APCG-leerlingen wat meer leerwinst; bij andere juist minder. Bij rekenen/wiskunde is de leerwinst van APCG-leerlingen steeds lager: behalve bij vmbo-kader scoren ze aan het eind van leerjaar 1 zelfs lager dan aan het begin.

5.3 Beginniveau en leerwinst naar 'voordeel van de twijfel'

Sommige scholen geven in hun toelatingsbeleid meer leerlingen 'het voordeel van de twijfel' dan andere scholen: zij laten leerlingen toe die een lager advies hebben dan de afdeling waarin zij beginnen, of een lagere score op de eindtoets basisonderwijs van Cito dan gebruikelijk is voor die afdeling. De vraag is hoe beginniveau en leerwinst van zulke leerlingen zich verhouden tot die van de overige leerlingen.

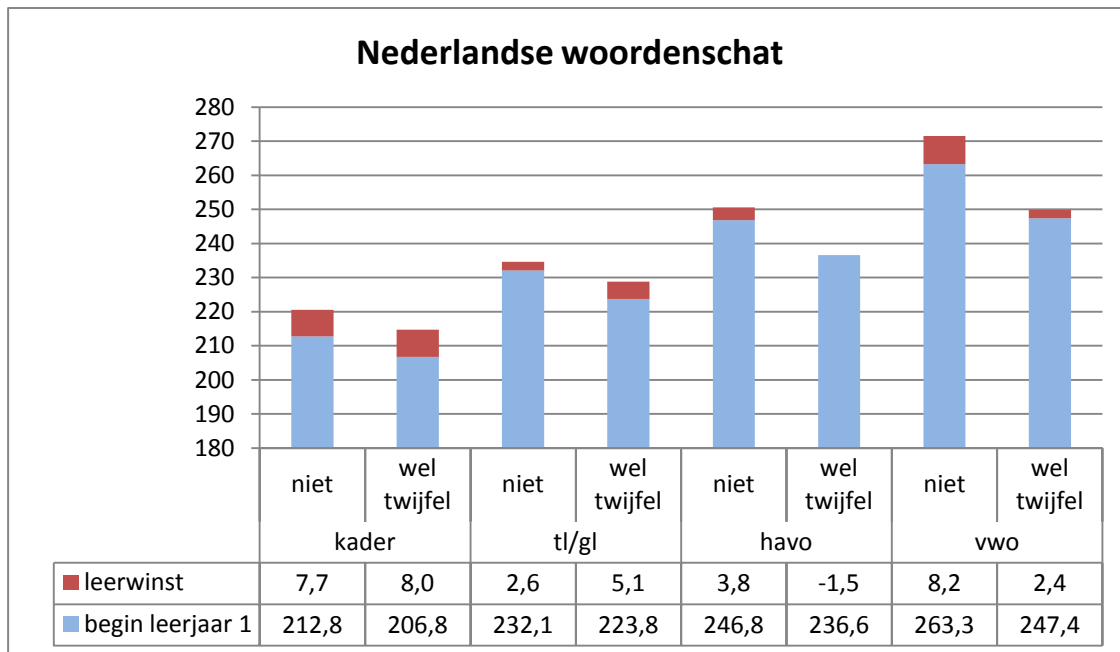
We zeggen dat een leerling het voordeel van de twijfel heeft gehad als het advies lager was dan het schooltype van instroom en/of de score op de Cito-eindtoets en/of de score op de NIO lager was dan gebruikelijk bij dit schooltype. Omdat er te weinig gegevens zijn over adviezen voor praktijkonderwijs (pro) en pro-leerlingen meestal geen Cito-eindtoets doen, kunnen we bij vmbo-basis niet goed bepalen wie er 'voordeel van de twijfel' heeft gekregen. We beperken ons daarom hier tot de overige vier onderwijsniveaus.

Figuur 5.5 Instroomniveau en leerwinst eerste leerjaar bij Nederlandse leesvaardigheid; naar 'voordeel van de twijfel'



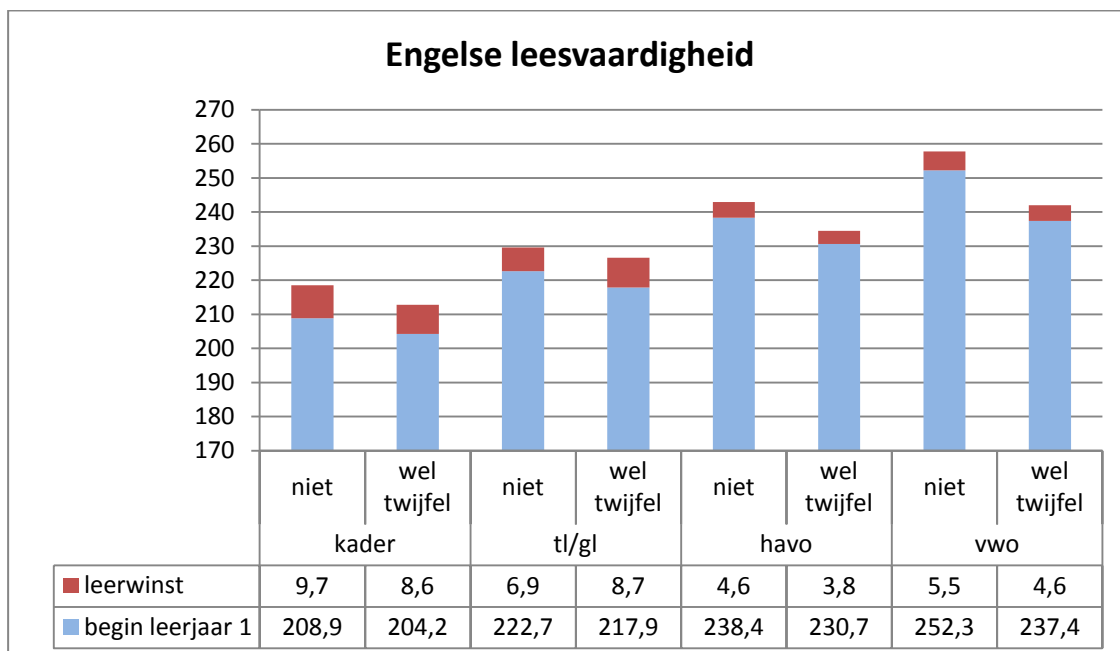
Bij Nederlandse leesvaardigheid zijn de beginscores van leerlingen met 'voordeel van de twijfel' duidelijk lager dan die van de overige leerlingen (effectgroottes van 0.3 bij kader tot 0.8 bij vwo). Verder boeken ze, behalve bij kader, ook minder leerwinst.

Figuur 5.6 Instroomniveau en leerwinst eerste leerjaar bij Nederlandse woordenschat; naar 'voordeel van de twijfel'



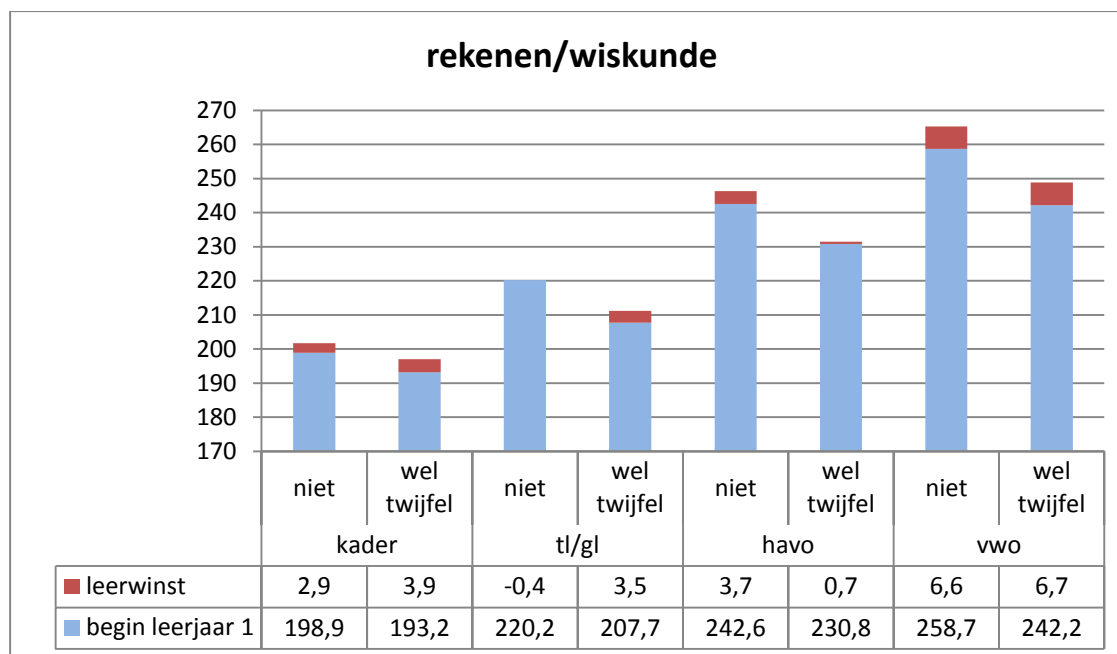
Ook bij Nederlandse woordenschat zijn de beginscores van leerlingen die 'het voordeel van de twijfel' hebben gekregen lager dan die van de overige leerlingen (effectgroottes van 0.3 tot 0.6). Bij beide vmbo-typen behalen ze vervolgens wel wat meer leerwinst, maar bij havo en vwo weer minder dan de overige leerlingen.

Figuur 5.7 Instroomniveau en leerwinst eerste leerjaar bij Engelse leesvaardigheid; naar 'voordeel van de twijfel'



Bij Engelse leesvaardigheid beginnen de leerlingen met 'voordeel van de twijfel' lager. Dat geldt in het bijzonder voor het vwo (effectgrootte 0.6). De daarna behaalde leerwinsten verschillen niet sterk tussen beide groepen leerlingen (alle effectgroottes kleiner dan 0.1).

Figuur 5.8 Instroomniveau en leerwinst eerste leerjaar bij rekenen/wiskunde; naar 'voordeel van de twijfel'



Ook bij rekenen/wiskunde is de start van leerlingen met 'voordeel van de twijfel' bij alle onderwijstypen lager dan die van overige leerlingen. De verschillen zijn het kleinst bij kader (effectgrootte 0.3), maar bij de andere onderwijstype fors (effectgroottes van 0.7 tot 0.9). Het beeld bij de leerwinst wisselt nogal: bij kader en vwo zijn er maar kleine verschillen, bij havo blijft de leerwinst van leerlingen met 'voordeel van de twijfel' achter, maar bij tl/gl boeken ze juist meer leerwinst (beide verschillen hebben een effectgrootte van rond de 0.25).

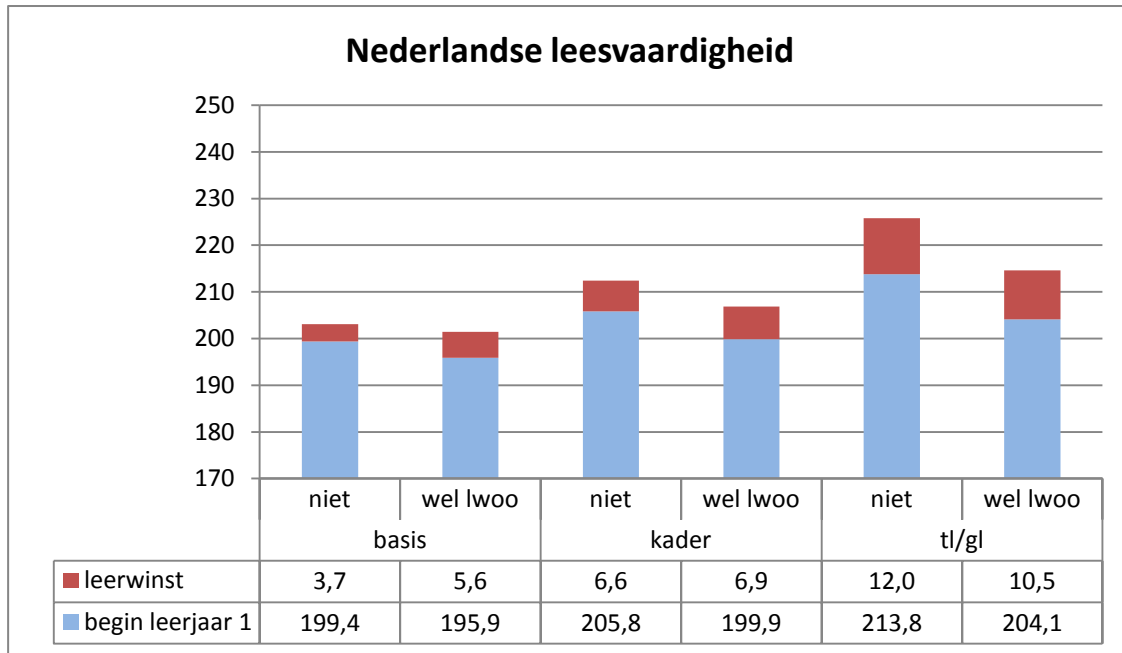
Samenvattend

Leerlingen die 'het voordeel van de twijfel' hebben gekregen stromen in met, zoals verwacht kon worden, lagere beginscores dan overige leerlingen. De leerwinsten laten echter een wisselend beeld zien per vaardigheid en per onderwijstype: soms blijven de leerlingen met 'voordeel van de twijfel' (ver) achter, soms halen ze juist meer leerwinst en soms is er weinig verschil.

5.4 Beginniveau en leerwinst naar leerwegondersteuning

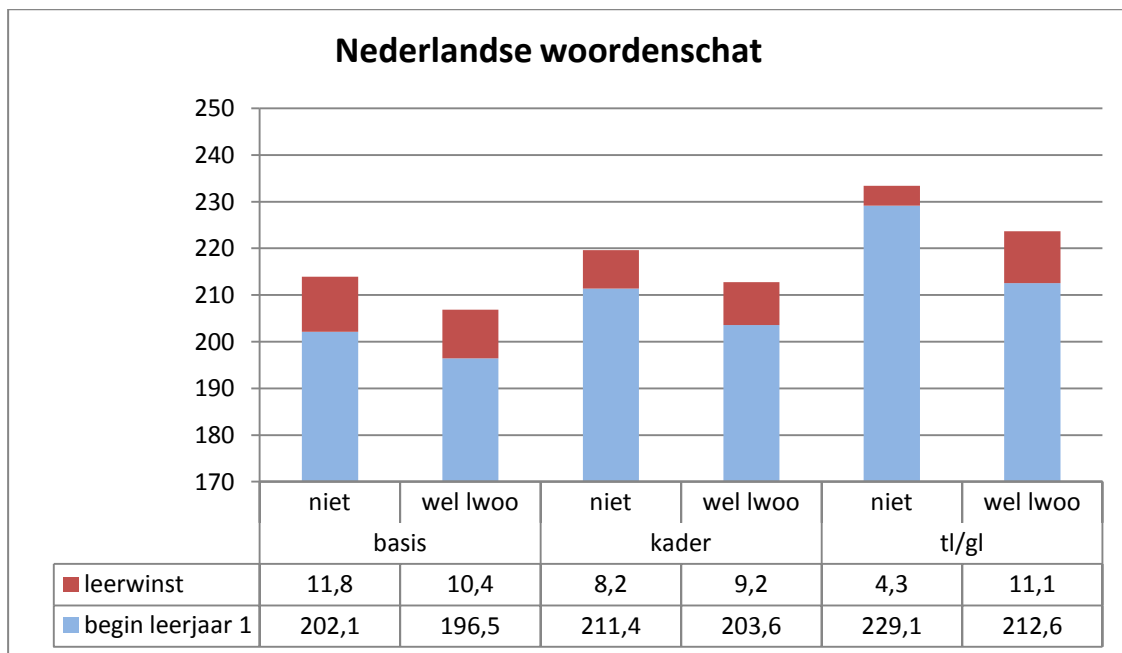
Binnen het vmbo zijn er leerlingen die in aanmerking komen voor leerwegondersteuning (lwoo). Ook hier gaan we weer na in hoeverre het instroomniveau en de leerwinst in het eerste leerjaar bij deze leerling verschillen met die van de overige leerlingen zonder lwoo.

Figuur 5.9 Instroomniveau en leerwinst eerste leerjaar bij Nederlandse leesvaardigheid; naar leerwegondersteuning



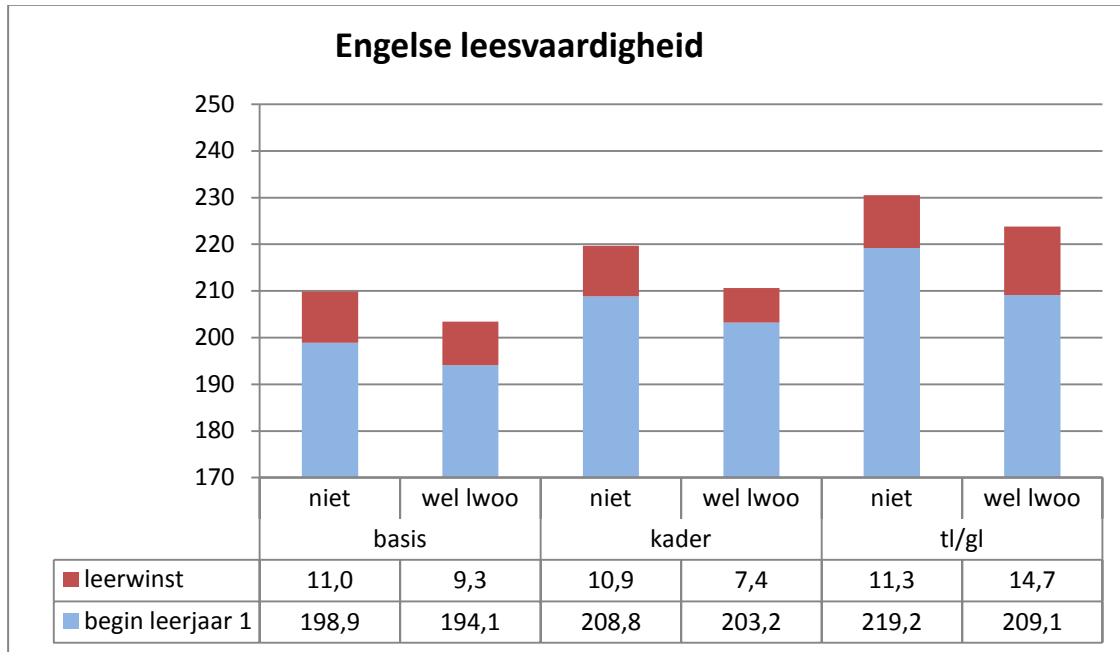
Bij alle drie typen vmbo komen de leerlingen met leerwegondersteuning binnen met lagere scores dan de overige leerlingen. De effectgroottes van het verschil lopen op van 0.3 bij basis, via 0.5 bij kader tot 0.8 bij tl/gl. Bij de leerwinsten is het beeld gunstiger: bij basis en kader lopen de lwoo-ers iets in, maar bij tl/gl is ook de leerwinst van de lwoo-ers wat lager.

Figuur 5.10 Instroomniveau en leerwinst eerste leerjaar bij Nederlandse woordenschat; naar leerwegondersteuning



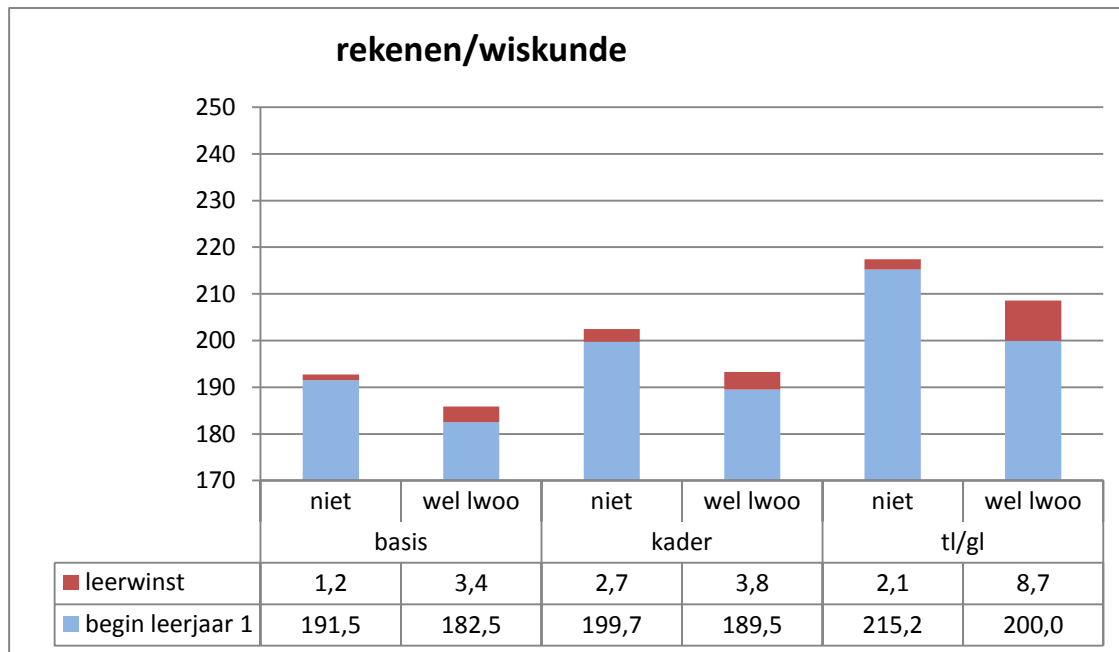
Ook bij Nederlandse woordenschat stromen lwoo-leerlingen in met lagere scores dan de overige leerlingen. De verschillen zijn opnieuw het grootst bij tl/gl (effectgrootte 0.7). De leerwinsten van lwoo-ers bij kader en basis verschillen niet veel van die van de overige leerlingen. Bij tl/gl lopen leerlingen met lwoo een flink stuk in op de overige leerlingen: ze behalen duidelijk meer leerwinst (effectgrootte 0.3).

Figuur 5.11 Instroomniveau en leerwinst eerste leerjaar bij Engelse leesvaardigheid; naar leerwegondersteuning



Opnieuw beginnen bij Engels de lwoo-leerlingen lager dan de overige leerlingen. Bij basis en kader zijn ook hun leerwinsten wat lager; bij tl/gl behalen ze juist wat meer leerwinst. De verschillen in leerwinst zijn echter steeds bescheiden (effectgrootte maximaal 0.15).

Figuur 5.12 Instroomniveau en leerwinst eerste leerjaar bij rekenen/wiskunde; naar leerwegondersteuning



Bij rekenen/wiskunde tenslotte starten de lwoo-ers ook weer op een flink lager niveau dan de overige leerlingen (effectgroottes van 0.5 bij basis tot 0.8 bij tl/gl). Daarna behalen ze overal wat meer leerwinst; vooral bij tl/gl is het verschil met de overige leerlingen aanzienlijk (effectgrootte 0.4).

Samenvattend

Leerlingen met een beschikking voor leerwegondersteuning starten, niet onverwacht, bij alle drie vmbo-typen en bij alle onderzochte domeinen, met een achterstand ten opzichte van de overige leerlingen. Vooral bij vmbo-tl/gl is het verschil vaak fors. Tegelijk behalen deze leerlingen juist bij tl/gl vaak meer leerwinst in het eerste leerjaar (uitzondering is Nederlandse leesvaardigheid). Het lijkt er daarmee op dat de tl/gl-afdelingen in de pilot er in slagen de leerlingen een deel van hun achterstand te laten inlopen. Bij basis en kader zijn de verschillen in leerwinst doorgaans kleiner en soms ook wisselend.

5.5 Verschillen tussen scholen en cohorten

In de vorige paragrafen is geïllustreerd welke verschillen er tussen leerlingen met diverse (achtergrond)kenmerken bestaan in instroomcores én in leerwinst in het eerste leerjaar. Bij het bepalen van de toegevoegde waarde van scholen wordt geprobeerd te corrigeren voor verschillen tussen scholen in zulke kenmerken van hun leerlingen. Bij sommige leerlingen blijkt het immers makkelijker te zijn om vooruitgang te boeken dan bij andere leerlingen. Bij leerlingen met 'ongunstige' achtergrondkenmerken zullen de verwachtingen omtrent de te behalen leerwinst lager zijn dan bij andere leerlingen. En een school met veel leerlingen met 'ongunstige' achtergrondkenmerken zal meer moeite moeten doen om de landelijke normen, gebaseerd op 'doorsnee- leerlingen', te behalen dan een school met veel leerlingen met juist 'gunstige' achtergrondkenmerken.

Zoals eerder opgemerkt (paragraaf 5.1) lijkt het weinig zinvol, gezien het beperkte aantal scholen dat aan de pilots deelneemt en gezien het feit dat zij samen geen representatieve steekproef vormen, om deze scholen

onderling te gaan vergelijken op hun, via statistische modellen geschatte, toegevoegde waarde. Wel hebben we geprobeerd een schatting te krijgen van de omvang van de verschillen in onderwijsresultaten tussen de beperkte groep pilotscholen, voor en na statistische correctie voor enkele relevante leerlingkenmerken. Doel van deze analyses is allereerst om een beeld te krijgen van de omvang van de systematische verschillen tussen deze scholen. Pas wanneer dergelijke verschillen ook substantieel zijn, heeft het zin om na te gaan of de kwaliteit van (het onderwijs op) de scholen een verklaring kan bieden voor deze verschillen. Wanneer de onderwijsopbrengsten vooral individueel bepaald zijn (binnen iedere school zijn er ongeveer evenveel leerlingen die het beter en die het slechter doen) presteren alle scholen ongeveer even goed.

Daarnaast leveren de analyses ook informatie op over de stabiliteit van de gevonden schoolverschillen over de jaren. Doordat we over gegevens van meerdere instroomcohorten beschikken, kan zowel worden nagegaan in hoeverre er systematische verschillen tussen scholen bestaan als in hoeverre er sprake is van fluctuaties van cohort tot cohort binnen dezelfde school.

Voor de schattingen van verschillen tussen scholen en cohorten hebben we analyses uitgevoerd met een multilevel random effect model. Met zo'n model wordt onderzocht in hoeverre de verschillen tussen leerlingen in toetsscores samenhangen met hun eerdere prestaties en met hun achtergrondkenmerken en kan vervolgens worden nagegaan of de dan resterende ('onverklaarde') verschillen in toetsscores systematisch samenhangen met de school die de leerlingen bezoeken. In het model wordt er expliciet rekening mee gehouden dat leerlingen 'genest' zijn binnen een bepaalde school. Klassieke literatuur over deze modellen zijn de publicaties van Aitkin en Longford (1986), Raudenbush & Bryk (1986) en Goldstein (1987). Een aantal recente Nederlandse toepassingen zijn te vinden in Rekers-Mombarg e.a. (2014), Timmermans (2012) en Roeleveld e.a. (2011).

In het model zijn de leerlingen genest binnen scholen. Maar omdat we ook de verschillen tussen de opeenvolgende instroomcohorten wilden verdisconteren hebben we gebruik gemaakt van een drienniveau model: leerlingen, genest in cohorten, genest in scholen. Voor zulke analyses kwamen de scholen met Cito- en Diataal-toetsen in aanmerking. We hebben gekozen voor de toetsen, waarbij we over de meeste gegevens beschikken, namelijk de toetsen van Cito.

Voor elk van de vier vaardigheidsscores die we binnen de pilot gebruiken, hebben we analyses uitgevoerd op de scores van leerlingen bij Toets 0 en bij Toets 1. De analyses van Toets 0 geven informatie over het instroomniveau van de leerlingen in de school. Bij de analyses van Toets 1, waarbij rekening wordt gehouden met de eerdere score op Toets 0, krijgen we informatie over de vooruitgang van de leerlingen in het eerste schooljaar. Bij beide typen analyses hebben we, als gezegd, onderzocht welke systematische verschillen er na controle voor het schooltype en enkele achtergrondkenmerken van de leerlingen overblijven tussen cohorten binnen eenzelfde school en tussen de pilotscholen onderling.

Idealiter zouden we in de analyses ook willen controleren voor verschillen tussen leerlingen in het opleidingsniveau van hun ouders. Uit tal van onderzoek blijkt immers dat dat een belangrijke voorspeller is van succes in het onderwijs (zie ook: Onderwijsraad 2013, 2014). Maar scholen voor voortgezet onderwijs beschikken niet standaard over gegevens met betrekking tot het ouderlijke onderwijsniveau en we kunnen slechts gebruik maken van een beperkt aantal andere achtergrondkenmerken, namelijk:¹⁸

- geslacht;
- woonachtig in een APCG-gebied;
- advies (geschaald van 1 = praktijkonderwijs, 2 = vmbo-basis tot 6 = vwo);
- indicatie voor 'voordeel van de twijfel' (zie paragraaf 5.3).

¹⁸ We gebruiken hier geen leerwegondersteuning omdat dat alleen bij vmbo-scholen relevant is.

Zoals gezegd hebben we ook verschillen tussen instroomcohorten op eenzelfde school geanalyseerd. Niet alle scholen hebben de toetsen de afgelopen vier jaar ook gebruikt, zodat het aantal cohorten wisselt per school. In totaal beschikten we voor de analyses over de gegevens van 3.770 leerlingen in 38 instroomcohorten op 13 scholen. De analyses zijn uitgevoerd met MIWin (Rasbash e.a., 2005).

Uitkomsten

De pilotscholen bieden onderwijs aan van vmbo-basis tot aan vwo. En uiteraard bestaan er dan grote verschillen in de toetscores die de leerlingen op die scholen behalen. De analyses van de behaalde scores op de verschillende onderdelen van de Cito-toetsen laten zien dat het onderwijstype en de achtergronden van de leerlingen een groot deel van deze verschillen tussen leerlingen verklaren. Van de resterende onverklaarde verschillen in toetsscores blijkt maar een klein deel systematisch samen te hangen met de school die de leerlingen bezoeken: in de orde van enkele procenten variantie. Verschillen tussen deze scholen in toegevoegde waarde zullen alleen betrekking hebben op deze beperkte hoeveelheid onverklaarde variantie.

Verder blijkt dat er ook verschillen bestaan tussen de opeenvolgende cohorten binnen de scholen; deze verschillen zijn van eenzelfde orde als die tussen de scholen. Dat betekent dat op dezelfde school de leerlingen het in het ene jaar wat beter doen en in het andere jaar wat slechter. En dat, voor zover een school in een bepaald jaar meer toegevoegde waarde zou laten zien dan de andere scholen, dat bij een volgend jaar weer niet zo hoeft te zijn.

Zulke verschillen tussen cohorten kunnen natuurlijk te maken hebben met veranderingen op de scholen, zoals wijzigingen in leerlinginstroom, verandering van leerkracht of aanpak en dergelijke. Bij de bespreking van de schoolrapporten in deze pilot konden enkele scholen veranderingen tussen cohorten ook goed duiden. Maar de verschillen kunnen ook meer 'toevallige' fluctuaties zijn. In wetenschappelijk onderzoek is al eerder gebleken dat schooleffecten niet heel stabiel in de tijd zijn (Roeleveld, 1994; Bosker & Luyten, 2000; Raudenbush, 2004).

Zoals gezegd hebben de analyses de nodige beperkingen. Er zijn weliswaar gegevens van veel leerlingen, maar slechts van een klein aantal scholen, die ook niet representatief zijn voor alle scholen voor voortgezet onderwijs in Nederland. Door het kleine aantal zullen de onderlinge verschillen ook niet snel statistisch significant zijn. Een verdere beperking is dat op lang niet alle onderzochte scholen ook steeds gegevens van vier opeenvolgende cohorten leerlingen beschikbaar zijn. Ook is de statistische controle voor verschillen in leerlingpubliek beperkt. Als gezegd zou vooral het opleidingsniveau van de ouders een belangrijke verbetering kunnen opleveren. Het is goed mogelijk dat bij een betere controle voor verschillen in leerlingpubliek er grotere onderlinge verschillen tussen de scholen worden gevonden.

Voor een meer gefundeerde analyse van de toegevoegde waarde van scholen voor voortgezet onderwijs zouden meer gegevens beschikbaar moeten zijn, van een groter aantal scholen, representatief voor het hele voortgezet onderwijs. Daarbij zouden ook betere gegevens over de sociale en culturele achtergrond van de leerlingen verzameld moeten worden, bijvoorbeeld het opleidingsniveau van de ouders. Onze uitkomsten laten verder zien dat zulke analyses in elk geval meer dan één leerlingcohort moeten omvatten.

5.6 Slotbeschouwing

Scholen voor voortgezet onderwijs verschillen onderling sterk in het soort leerlingen dat de school bezoekt. Dat heeft natuurlijk vooral te maken met de verschillende niveaus van voortgezet onderwijs, van vmbo-basis tot aan gymnasium, die de scholen aanbieden. Maar ook bij scholen met hetzelfde onderwijsniveau kunnen er grote verschillen bestaan in de achtergronden van hun leerlingen. Bij het bepalen van de toegevoegde waarde

van een school wordt getracht de bijdrage van de school aan de ontwikkeling van leerlingen te isoleren van zulke verschillen tussen hun leerlingen en van verschillen in de context van de school.

In dit hoofdstuk hebben we laten zien dat leerlingen met verschillende achtergrondkenmerken gemiddeld ook verschillen in zowel hun aanvangsniveau als in de leerwinst die zij boeken. Leerlingen uit APCG-gebieden, leerlingen met leerwegondersteuning en leerlingen, die ondanks een lager advies of eindtoetsscore toch zijn toegelaten, starten met lagere toetsscores dan de overige leerlingen. Voor de leerwinst die deze leerlingen behalen is het beeld minder eenduidig: soms lopen ze een deel van hun aanvankelijke achterstand in, soms blijven ze juist nog verder achter in het eerste leerjaar.

Er is dus alle reden om bij het beoordelen van de opbrengsten van scholen rekening te houden met zulke verschillen tussen die scholen in hun leerlingenpopulatie. Daarvoor zijn echter geavanceerde statistische modellen nodig, die bovendien gegevens vereisen van een groot aantal landelijk representatieve scholen. Voor een afzonderlijke school is het daarom niet goed mogelijk om zelf precies de eigen toegevoegde waarde, in vergelijking met die van andere scholen, te bepalen.

Vanwege het beperkt aantal deelnemende scholen, met verschillende onderwijsniveaus en met uiteenlopende toetsen en andere instrumenten, is het in deze pilot niet goed mogelijk geweest om de toegevoegde waarde van afzonderlijke scholen op een betrouwbare manier te bepalen. Uit verkennende analyses is wel gebleken dat, na correctie voor het niveau van aangeboden onderwijs en voor enkele achtergrondkenmerken van de leerlingen, er slechts relatief kleine verschillen tussen de pilotscholen bestaan in hun onderwijsresultaten. Bovendien bleek dat er tussen opeenvolgende cohorten binnen dezelfde school verschillen van vergelijkbare grootte optreden.

6 Ervaringen van de scholen

In dit hoofdstuk gaan we in op de ervaringen van de scholen in de pilot. Welke meerwaarde heeft de deelname voor de scholen gehad, wat zijn de opbrengsten voor de scholen en hoe zijn de scholen van plan om verder te gaan met leerwinst en – meer in het algemeen – opbrengstgericht werken? De hier gerapporteerde ervaringen van de scholen zijn gebaseerd op de evaluatie met de pilotscholen, die gelijktijdig met de bespreking van de tweede schoolrapportage plaats vond. Ook hebben we gebruik gemaakt van de bevindingen van de gezamenlijke slotbijeenkomst van de pilot. Ter illustratie van de ervaringen hebben we in dit hoofdstuk een aantal citaten van pilotscholen opgenomen.

6.1 Meerwaarde pilot

Opbrengsten/verwachtingen pilot algemeen

In hoofdstuk 2 gaven we aan dat scholen verschillende verwachtingen hadden bij de pilot. Bij een deel ging het om meer zicht te krijgen op de rol van achtergrondkenmerken in de ontwikkeling en resultaten van de leerlingen. Andere scholen hoopten met behulp van de in de pilot ontwikkelde instrumenten op het gebied van leerwinst en toegevoegde waarde meer zicht te krijgen op de ontwikkeling van de cognitieve prestaties of het onderwijsrendement en meer handvatten te krijgen om het onderwijs te verbeteren. Een aantal scholen wilde graag ook zicht krijgen op de niet-cognitieve ontwikkeling van hun leerlingen.

Uit de evaluatie blijkt dat de verwachtingen van de scholen op hoofdlijnen zijn uitgekomen. De meeste scholen zien een meerwaarde in de deelname aan de pilot door het inzicht dat ze hebben gekregen in de leerwinst van hun leerlingen en het nut van het afnemen en (scherper) analyseren van (methode-onafhankelijke) toetsen. In de loop van het project zijn er echter soms personele wisselingen geweest op de scholen, waardoor een nieuwe directeur met andere (of geen) verwachtingen van de pilot instroomde. Daarnaast hadden sommige scholen specifieke wensen of verwachtingen die niet altijd konden worden ingelost in het kader of tijdsbestek van de pilot. Soms bracht de pilot voor de scholen ook andere inzichten, bijvoorbeeld dat er verbeteringen mogelijk zijn in de administratie van gegevens. Een aantal scholen geeft aan dat de pilot te kort was om aan alle verwachtingen te voldoen. Voor een deel van de scholen was de pilot ook onderdeel van een breder proces rondom opbrengstgericht werken. Afhankelijk van hoe ver de school hier al in was, heeft de pilot hier in meer of mindere mate aan bijgedragen. Een aantal scholen geeft aan het jammer te vinden dat het niet gelukt is om een maat voor de toegevoegde waarde van de school te ontwikkelen.

“Door de pilot is de school er achter gekomen dat de administratie mankementen vertoont. Dit is voor de school een leerpunt geweest. Doordat bepaalde zaken niet goed waren geregistreerd of uit het systeem konden worden geëxporteerd heeft de pilot minder inzichten opgeleverd dan gehoopt.”

“De pilot heeft binnen de school tot gevolg gehad dat er ook meer het gesprek over de resultaten wordt aangegaan met de onderbouwdocenten i.p.v. alleen met de bovenbouwdocenten. Ook is er nog sterker ingezet op woordenschat en rekenen, o.a. door specifiek beleid hierop voor leerjaar 1 t/m 4. Er is meer grip, inzicht en bewustzijn in de onderbouw, doordat trends meer zichtbaar zijn.”

“De pilot heeft bijgedragen aan bewustwording bij de directeur en bij de jaarlaagcoördinatoren. Zij zijn op het spoor gezet van opbrengstgericht werken en op de mogelijkheden die leerwinst hiervoor bieden. Zij zien nu in dat het waardevol is om op een andere manier, via cijfers, naar de school te kijken om zo zicht te krijgen op verbeteringen.”

Schoolrapportages

De scholen hebben gedurende de pilot twee keer een eigen schoolrapportage ontvangen. De inhoud van de rapportages was voor de cognitieve vakken afhankelijk van het gebruikte toetsystemen (of toetssystemen als er meerdere werden gebruikt), de beschikbaarheid van (meerjarige) toetsgegevens voor het schooltype en leerjaren waar de school voor had gekozen en van de beschikbare administratieve gegevens van de leerlingen. Daarnaast kregen scholen die ervoor hadden gekozen om ook de niet-cognitieve ontwikkeling te willen betrekken in de pilot ook de resultaten hierop in de rapportage teruggekoppeld.

De rapportages gaven voor de meeste scholen op een of meerdere onderdelen een bevestiging van het beeld dat de scholen zelf hadden, maar boden ook aanvullende informatie boven op de analyses die ze zelf hadden gemaakt. De meerwaarde had vooral betrekking op inzicht in de leerwinst van een cohort leerlingen over meerdere jaren, een vergelijking tussen verschillende cohorten leerlingen en een vergelijking van de prestaties en leerwinst van de eigen leerlingen met landelijke normen. Ook het feit dat er verschillende vakken naast elkaar worden gezet was voor veel scholen een eye-opener. Zo werd bijvoorbeeld zichtbaar dat het instroomniveau van leerlingen niet voor alle vakken hetzelfde hoeft te zijn en dat er verschillen kunnen zijn in de leerwinst die bij verschillende (deel)vakken wordt geboekt (bijv. meer bij begrijpend lezen dan bij woordenschat). Scholen maken zelf vooral analyses op klasniveau en leerlingniveau en er wordt ook vaak minder gekeken naar verschillen in ontwikkeling tussen vakken (bijv. rekenen vs. begrijpend lezen).

De informatie die in de rapportages is opgenomen over de achtergrondkenmerken van leerlingen (bijv. wat betreft APCG-gebied, sekse, lwoo, op- en afstroom) in relatie tot toetsresultaten is voor veel scholen ook interessante extra informatie, die ze zelf vaak niet in verband brengen met de resultaten.

De pilotrapportages hebben in het algemeen ook vaak tot gevolg gehad dat de opbrengsten meer binnen de school en met docenten besproken zijn en heeft voor een deel van de scholen bijgedragen tot het nemen van concrete acties.

“Het schoolrapport heeft zeker meerwaarde ten opzichte van de eigen gegevens. Van de Cito-resultaten worden alleen op leerling- en groepsrapportages uitgedraaid. Er worden bij ons geen rapporten gegenereerd waarin leerjaren en cohorten vergeleken kunnen worden.”

“De meerwaarde van de pilotrapportage ten opzichte van de eigen rapportage is dat deze een beeld geeft van de resultaten op een wat langere termijn en een vergelijking tussen cohorten. Dit geeft ook meer druk om er echt iets mee te doen en er acties op in te zetten. “Niet analyseren en het daarbij laten”. Het is moeilijk zelf een dergelijke rapportage te maken, met name wat betreft de leerwinst en de landelijke vergelijking. Door deze extra inzichten dragen de pilotrapportages bij aan het verbeteren van de school.”

“Het schoolrapport heeft meerwaarde ten opzichte van de analyses die nu gedaan worden op basis van de Cito-gegevens en de gegevens uit Magister. Deze gegevens zijn op dit moment nog niet aan elkaar gekoppeld en laten dus slechts delen van resultaten zien. Het schoolrapport biedt verder de mogelijkheid om te vergelijken met het landelijke gemiddelde en biedt extra uitsplitsingen naar geslacht, APCG, lwoo en de vertraging bij instroom.”

Voor scholen die zelf al uitgebreidere analyses maken, had de rapportage uiteraard minder meerwaarde. Ook was het voor enkele scholen die bepaalde toetssystemen (bijv. TOA) gebruiken jammer dat er geen landelijke vergelijking mogelijk was. Een aantal scholen gaf aan behoefte te hebben aan een vergelijking met scholen met een zelfde leerlingpopulatie of met scholen uit dezelfde regio.

Veel scholen geven aan de rapportage te hebben gedeeld met de relevante vaksecties (team- of afdelingsleiders en docenten). Daarbij gaf een deel wel aan dat er soms een vertaalslag nodig was voor met name docenten om de resultaten te kunnen interpreteren, omdat het kijken naar leerwinst nog nieuw voor ze is. De rapportages werden ook vaak gedeeld met het bestuur. Voor ouders werd de rapportage minder geschikt geacht.

6.2 Ondersteuning

Over het algemeen zijn de scholen tevreden over de ondersteuning die zij tijdens de pilot hebben gekregen. Deze ondersteuning bestond uit hulp bij de aanlevering van de gegevens, uitleg over de pilotrapporages aan betrokkenen op school, gezamenlijke bijeenkomsten met alle pilotscholen en – indien gewenst en mogelijk binnen de kaders van de pilot – specifieke ondersteuning op maat.

De organisatie rondom het aanleveren van de gegevens en het bespreken van de rapportages op school werden door vrijwel alle scholen als goed en zinvol beschouwd. Bij de bespreking van de rapportages hadden sommige scholen meer behoefte aan interpretatie en duiding van de gepresenteerde informatie dan andere scholen. Over de eerste twee gezamenlijke bijeenkomsten voor alle pilotscholen waren de meningen wat verdeeld.¹⁹ Sommige scholen vonden de bijeenkomsten heel nuttig en leerzaam, bijvoorbeeld vanwege de contacten en uitwisseling met andere scholen, maar andere scholen gaven aan hier minder aan te hebben gehad. De wisselende ervaringen met de eerste twee gezamenlijke bijeenkomsten hebben wellicht te maken met hoe ver de school al was op het gebied van opbrengstgericht werken met behulp van toetsgegevens of met de specifieke context van de school (achtergrond leerlingen/docenten, onderwijstype, regionale context). Over het algemeen waren de scholen tevreden over de extra ondersteuning op maat die ze hadden ontvangen (zie hoofdstuk 2), al werd bij de workshops over opbrengstgericht werken door een paar scholen aangegeven dat deze voor sommige docenten wat ingewikkeld waren of nog te ver af stonden van hun dagelijkse lespraktijk.

“Vooral de besprekingen van de rapportages zijn als prettig en zinvol ervaren. Het administratieve deel van de pilot was erg goed georganiseerd, de belasting voor de school was daardoor heel beperkt. De gezamenlijke bijeenkomsten hebben niet zoveel opgeleverd voor de school, uit de interactie met andere scholen is niet veel gekomen en de verhalen/workshops waren te algemeen.”

“De school is erg blij met het bespreken van het schoolrapport. Het helpt het rapport beter te begrijpen. Daarnaast is de rapportbespreking een impuls om er op school meer mee te doen.”

“De contacten met andere scholen zijn altijd waardevol. De school is zelf al redelijk ver met opbrengstgericht werken, maar toch doe je door de contacten altijd ook nieuwe inzichten op.”

“Zowel de bespreking van de rapportages als de workshop opbrengstgericht werken worden door de school positief beoordeeld. Dit was goed, nuttig en stimulerend. We hebben veel geleerd over de interpretatie van data.”

¹⁹ De evaluatie op de pilotscholen vond plaats voor de gezamenlijke slotbijeenkomst. We hebben daarom geen uitgebreid beeld van de tevredenheid van de scholen over deze slotbijeenkomst. Wel weten we uit de plenaire afsluiting van de slotbijeenkomst dat de aanwezigheid van de slotbijeenkomst nuttig vonden.

“De workshop opbrengstgericht werken werd goed gewaardeerd. Voor sommige collegae was het wellicht wat te abstract, die hebben behoefte aan kennis en vaardigheden die ze direct de volgende dag in de klas kunnen gebruiken.”

6.3 Onderwerpen pilot

De pilots leerwinst en toegevoegde waarde werden uitgevoerd binnen de context van opbrengstgericht werken. Binnen dit kader is gekeken naar de bruikbaarheid van het concept van leerwinst voor het volgen van resultaten op cognitieve vakken (Nederlands begrijpend lezen, Nederlands woordenschat, Engelse leesvaardigheid, rekenen/wiskunde) en op niet-cognitief gebied (motivatie en burgerschap).

Opbrengstgericht werken

Aan het eind van de pilot hebben alle scholen aangegeven verder te willen gaan met opbrengstgericht werken. Voor de meeste scholen was de pilot één van de zaken die daaraan een impuls aan heeft gegeven. Zij waren al in meer of mindere mate bezig met opbrengstgericht werken. Hun deelname aan de pilots was daarin een (vervolg)stap. Enkele scholen werkten op het gebied van opbrengstgericht werken ook al samen met externe partijen, zoals het datateam van de Universiteit Twente of School aan Zet. Ook wordt in het onderwijsveld steeds meer gebruik gemaakt van managementtools van de gebruikte onderwijsinformatiesystemen in het kader van opbrengstgericht werken.

De meeste scholen geven aan dat ze de pilotrapportages bruikbaar vinden in het kader van opbrengstgericht werken. De pilot in het algemeen en de schoolrapportages in het bijzonder hebben bij de meeste scholen voor meer bewustzijn gezorgd over de bruikbaarheid van het analyseren van toetsresultaten bij opbrengstgericht werken. Daarnaast heeft de pilot ook het nut laten zien van het kijken naar leerwinst, omdat goede resultaten an sich niet alles zeggen: “Als docenten alleen maar kijken naar het absolute niveau (los van leerwinst) kan dat aanleiding geven om maar achterover te leunen.”

“Het rapport levert stof op tot discussie en laat zien welke perspectieven er zijn om data te gebruiken. De inhoud van het rapport roept ook nieuwe vragen op. Deels gaat het om vragen waarvoor meer analyses nodig zijn, bijvoorbeeld de vraag hoe doen leerlingen die hoger geplaatst worden dan hun advies het in hogere leerjaren? Deels gaat het om vragen die besproken moeten worden met de secties, bijvoorbeeld wat zijn mogelijke verklaringen voor de verschillen in eindexamenresultaten tussen de vakken Nederlands en Engels?”

“Ja, we kunnen het rapport gebruiken bij opbrengstgericht werken. Er lag bijvoorbeeld een tijd veel nadruk op Engels in de bovenbouw. Nu blijkt dat de scores in de onderbouw al tegenvallen, is de nadruk verlegd naar de onderbouw.”

“De school kan de rapportage gebruiken bij opbrengstgericht werken. Zo worden de uitkomsten van de rapportage gedeeld met de betreffende docenten om te laten zien hoe gescoord wordt ten opzichte van het landelijk gemiddelde en dat er mooie resultaten behaald worden op deze school.”

Aan de scholen is gevraagd wat zij nodig hebben om verder te gaan met opbrengstgericht werken. Scholen noemen verschillende aspecten die hierbij van belang zijn, zoals binnen de school/vaksecties een plan en doelen opstellen, zorgen voor een goed toetsbeleid, professionalisering van docenten (met name wat betreft analyseren en interpreteren van toetsresultaten en deze koppelen aan concrete acties en het verhogen van het databewustzijn). Een aantal scholen heeft ook goede ervaringen met de ondersteuning van overige externe partijen en wil hier graag mee verder gaan in de toekomst.

Cognitieve vakken

Aan de scholen die voor de cognitieve vakken hebben deelgenomen aan de pilot is aanvullend nog een aantal vragen gesteld over de omgang met en gebruik van de toetssystemen en -resultaten binnen de school. Hieruit komt naar voren dat de toegang tot het webportal van het betreffende systeem meestal op (midden-) managementniveau ligt (directie, taal-/of rekencoördinatoren, afdelings- of teamleiders). Deze personen beheren de gegevens en maken de benodigde rapportages, die verder binnen de teams worden verspreid en besproken. Daarbij wordt veel op individueel en klasniveau gekeken, analyses op vak- of schoolniveau komen minder vaak voor. Op een aanzienlijk deel van de scholen is het maken van rapportages op basis van de toetsresultaten binnen de systemen nog niet heel sterk verankerd en er wordt ook niet van alle mogelijkheden gebruik gemaakt. Door de pilot is er het besef gekomen dat er soms nog wel een slag kan worden geslagen in de kennis over en gebruik van de mogelijkheden van de toetssystemen, met name bij docenten.

“Binnen de school hebben de decaan, teamleiders en directie toegang tot het webportal van Cito. Docenten en mentoren hebben niet direct toegang. De rapportages worden door de teamleiders gemaakt en met de docenten besproken op de rapportvergaderingen. Er vindt ook een terugkoppeling plaats van de resultaten op de Cito-toetsen aan de ouders.”

“De afdelingsleiders hebben toegang tot het Cito Volgsysteem en draaien de rapportages uit voor de docenten (en directie). De rapportages worden op per leerjaar op leerling- en klasniveau uitgedraaid, geanalyseerd en besproken. De leerlingbegeleiders hebben in principe ook inzage in de resultaten, maar zij maken gebruik van de uitdraaien die worden aangeleverd door de afdelingsleiders. De betrokkenheid van de docenten bij de resultaten van de Cito-toetsen zou nog beter kunnen.”

Motivatie en burgerschap

Enkele scholen hebben in het kader van de pilot de vragenlijsten over motivatie en burgerschap afgenomen. Aan deze scholen is gevraagd in hoeverre de gebruikte vragenlijsten uit het cohortonderzoek COOL⁵⁻¹⁸ aansluiten bij wat zij willen meten. Twee scholen geven aan dat het voordeel van de COOL-vragenlijsten is dat er een landelijke vergelijking mogelijk is. De ander scholen hadden hier geen duidelijk zicht op.

“Wij hebben een eigen LVS dat juist deze aspecten meet; beter nog dan de aspecten die in het COOL-onderzoek worden meegenomen. Helaas is ons eigen instrument niet gebenchmarkt. De motivatievragenlijst uit het COOL-onderzoek geeft ons wel inzicht in de manier waarop leerlingen gemotiveerd zijn.”

“We vinden het prettig om met landelijk genormeerde vragenlijsten te werken, waarvan ook benchmarkgegevens bekend zijn.”

6.4 Conclusies en adviezen van de scholen

Uit de ervaringen van de pilotscholen komt naar voren dat men over het algemeen tevreden is over het verloop en de opbrengsten van de pilot. De pilot heeft tot meer bewustwording geleid van de concepten leerwinst, toegevoegde waarde en opbrengstgericht werken. De pilotscholen hebben op basis van hun ervaringen in de pilot nog wel een aantal tips en adviezen voor andere scholen, toetsleveranciers en voor de inspectie en het ministerie van OCW om opbrengstgericht werken nog verder te versterken. We geven de feedback van de scholen hier beknopt weer. De feedback varieert van algemene tips tot meer specifieke aandachtspunten en adviezen. Het gaat hierbij meestal om opmerkingen van een enkele school.

Voor andere scholen:

- Laat docenten (externe) scholing volgen op het gebied van opbrengstgericht werken.
- Leer van elkaar. Ga bij andere (voorbeeld)scholen langs. Een school geeft in dit kader aan dat het mooi zou zijn als er een overkoepelende organisatie zou zijn om dit soort uitwisselingen te organiseren.
- Gewoon doen en blijven doen. Opbrengstgericht werken vanuit de directie blijven agenderen en blijven werken aan bewustzijn.
- Vier successen en realiseer je dat het nooit klaar is.
- Stel bepaalde docenten vrij voor het rekenonderwijs.

Voor toetsmakers en -leveranciers:

- Er is behoefte aan landelijke diagnostische toetsen voor alle vakken en betere methode-afhankelijke toetsen.
- De koppeling tussen toetsgegevens en leerlingadministratie is bij sommige systemen nog een belemmerende factor.
- Het rapporteren over de referentieniveaus zou meer afgestemd moeten worden tussen de verschillende toetssystemen, zodat er een betere vergelijking mogelijk is.
- Sommige scholen hebben behoefte aan tools binnen het gebruikte toetssysteem die helpen bij het trekken van conclusies uit de resultaten.
- Meer afstemming op het 'schoolritme' bij de toetsafname, bijv. wat betreft de toetsperiode en de duur van de toetsafname (niet meer dan 1 lesuur).
- Gebruiksvriendelijkheid: technische problemen werken vaak ontmoedigend voor docenten die het toch al druk hebben.
- Er wordt aandacht gevraagd voor de 'taligheid' van veel rekentoetsen, bijvoorbeeld die van Cito, wat problematisch kan zijn voor taalzwakke leerlingen. Hier zou bij de ontwikkeling van toetsen nog meer rekening mee kunnen worden gehouden.
- Veel toetssystemen gaan nu nog maar tot de onderbouw, een uitbreiding naar de bovenbouw is gewenst, zodat scholen de ontwikkeling van de leerlingen ook in de hogere leerjaren op dezelfde wijze kunnen blijven volgen.

"Er zijn nu alleen methode-onafhankelijke toetsen voor de kernvakken in de onderbouw. Om leerlingen gedurende hun gehele vo-periode te kunnen volgen zouden er ook toetsen voor de bovenbouw moeten komen."

Voor OCW:

- Een deel van de scholen geeft aan dat ze graag zouden zien dat de pilot wordt voortgezet. Sommige scholen zouden graag jaarlijks een schoolrapportage blijven ontvangen, terwijl andere behoefte hebben aan een maat voor de toegevoegde waarde van de school.
- Geef scholen de tijd en ruimte om veranderingen door te voeren en te bestendigen.
- Er wordt aandacht gevraagd voor de haalbaarheid van de referentieniveaus, met name door vmbo-basis, waarbij voor veel leerlingen niveau 2F niet haalbaar is. Leerlingen zouden daar niet teveel op afgerekend moeten worden, het gaat ook om de leerwinst die ze behalen.

"De pilot leerwinst en toegevoegde waarde zou (door OCW) moeten worden voortgezet als blijkt dat toegevoegde waarde nog niet gemeten kan worden. Het kan helpen als er meer en langer gegevens worden verzameld over meerdere cohorten. OCW moet daarin een partner zijn van het onderwijsveld."

Voor de inspectie:

- Enkele scholen geven aan dat de inspectie meer rekening zou moeten houden met bijzondere omstandigheden op sommige scholen, waardoor het soms nodig is om buiten de bestaande kaders te kijken. Maar er wordt door andere scholen ook waardering uitgesproken voor het feit dat de inspectie hierover meedenkt.
- Scholen geven ook aan behoefte te hebben aan een ruimere definitie van opbrengsten. De huidige focus ligt volgens hen op cognitieve kernvakken en toetsen.
- Scholen staan over het algemeen positief tegenover het concept van toegevoegde waarde, maar zijn wel huiverig voor het ontstaan van een afrekencultuur.

“Het is prettig te merken dat er goed wordt meegedacht door de inspectie en dat ook de scholen die inzetten op de brede ontwikkeling van leerlingen gehoord worden.”

“Voorkomen moet worden dat er een afrekencultuur ontstaat rondom het begrip toegevoegde waarde, omdat scholen anders er alleen aan zullen werken dat ze er zo goed mogelijk uitkomen, in plaats van dat ze inzicht in hun toegevoegde waarde gebruiken om hun onderwijs te verbeteren.”

7 Conclusies en aanbevelingen

Deze pilot leert ons dat het goed mogelijk is voor scholen om zelf leerwinst inzichtelijk te maken. Voor de toegevoegde waarde van scholen ligt dat anders. Wij concluderen dat scholen, in het kader van opbrengstgericht werken, meer kunnen doen met de data die ze al tot hun beschikking hebben. Wij bevelen aan dat de Inspectie van het Onderwijs dit stimuleert door de wijze waarop de scholen hun 'leerwinst' inzichtelijk maken, mee te nemen in de beoordeling van de kwaliteitszorg van scholen.

De onderzoeksvragen voor de pilot, die wij in hoofdstuk 1 gepresenteerd hebben, luiden als volgt:

1. Welke mogelijkheden zijn er om de leerwinst van leerlingen in kaart te brengen en te meten?
2. Welke mogelijkheden zijn er om op basis van de informatie die de school zelf heeft een indicatie van de toegevoegde waarde te verkrijgen en mogelijk tot een gestandaardiseerde methode te komen?
3. Welke rol kan het meten van de leerwinst van leerlingen en het mogelijk meten van een toegevoegde waarde spelen bij de verdere ontwikkeling van het opbrengstgericht werken?
4. Welke ontwikkelingen binnen de pilots en welke producten uit de pilots kunnen worden ingezet bij het verder ontwikkelen van de opbrengstsystematiek en de ontwikkelingen ten aanzien van het (nieuwe) waarderingskader van de inspectie?

In dit slothoofdstuk lichten we de conclusies en aanbevelingen per onderzoeksvraag toe.

7.1 Welke mogelijkheden zijn er om de leerwinst van leerlingen in kaart te brengen en te meten?

Leerwinst inzichtelijk maken

In deze pilot is het goed gelukt om leerwinst inzichtelijk te maken met behulp van data waarover vo-scholen beschikken. Op verschillende aspecten zoals leesvaardigheid (Nederlands en Engels), woordenschat, rekenen/wiskunde zijn opeenvolgende metingen op schoolniveau vergeleken met een norm (het landelijk gemiddelde). De resultaten zijn in tabellen en grafieken gepresenteerd. Hiermee wordt inzichtelijk gemaakt hoe de groei (leerwinst) op schoolniveau zich verhoudt tot het landelijk gemiddelde. Ook zijn vragenlijsten voor motivatie en burgerschap afgenomen. Voor deze aspecten is het niet goed gelukt om ontwikkelingen ('leerwinst') vast te stellen omdat er maar weinig longitudinale metingen bij dezelfde leerlingen zijn uitgevoerd. Bovendien ontbreekt vooralsnog een landelijke norm om deze ontwikkeling vervolgens te beoordelen. Dit neemt niet weg dat het ook hiervoor mogelijk moet zijn voor scholen om zelf de 'leerwinst' inzichtelijk te maken, mits de metingen meerdere malen bij dezelfde groep leerlingen worden afgenomen en er een landelijke norm beschikbaar is.

Uit deze pilot blijkt dat scholen in de praktijk verschillende instrumenten gebruiken waarmee leerwinst kan worden vastgesteld. Dit betreft vooral niveaubepalingen voor Nederlands, Engels en rekenen in de onderbouw. De toetsen zijn genormeerd aan de hand van landelijke gemiddelden en gerelateerd aan de referentieniveaus voor taal en rekenen of de ERK-niveaus (Europese referentiekader) voor Engels. Voor niet-cognitieve vaardigheden is het niet gebruikelijk dat scholen genormeerde metingen uitvoeren, hier is nog ruimte voor ontwikkeling. Bij een deel van de pilotscholen is er duidelijk behoefte aan betere instrumenten om de beheersing van niet-cognitieve vaardigheden inzichtelijk te maken. Ook voor andere vakken dan de kernvakken en voor de bovenbouw zijn er geen landelijk genormeerde toetsen af te nemen anders dan het centraal eindexamen.

Criteria voor het meten van leerwinst

Voor het inzichtelijk maken van leerwinst is het nodig om een cohortanalyse uit te voeren. Een meting moet meerdere malen bij dezelfde groep leerlingen worden uitgevoerd. Dus meerdere metingen waarbij de score op dezelfde schaal wordt uitgedrukt. Gebruikelijk zijn vaak drie of vier meetmomenten in de eerste drie leerjaren. Op de pilotscholen gebeurde dit niet altijd, met name niet voor de vragenlijsten betreffende motivatie en burgerschapscompetenties. Daarbij is het van belang dat álle leerlingen meedoen aan de meting. Als leerlingen geselecteerd worden (bijvoorbeeld door de zwakste of de sterkste leerlingen uit te sluiten), vertekent dit het resultaat op schoolniveau. Om onderling vergelijkbaar te zijn is het ook van belang dat de toetsen op de scholen op een vergelijkbaar tijdstip en onder vergelijkbare condities (geen hulp) worden afgenomen.

Om vervolgens een uitspraak te kunnen doen over de gerealiseerde leerwinst moet er een norm zijn. Vaak gaat het hier om het landelijk gemiddelde voor een bepaalde onderwijssoort (basis, kader, theoretisch, havo of vwo). De toetsen hebben dus een landelijke norm waarmee scholen zichzelf en hun leerlingen met het landelijk gemiddelde van andere scholen kunnen vergelijken. Voor toetsen, die alleen de beheersing van de referentieniveaus weerspiegelen, is leerwinst veel minder goed aan te geven. Voor bijvoorbeeld een leerling die op twee momenten op niveau 2F scoort, is niet duidelijk hoeveel leerwinst deze heeft geboekt. Dat kan bijvoorbeeld best veel zijn als de score de eerste keer maar net boven niveau 1F lag en de tweede keer net onder 3F.

Een andere eis aan de meetinstrumenten (meestal toetsen of vragenlijsten) is dat zij dekkend zijn voor datgene wat gemeten moet worden. Belangrijk hierbij is dat in het voortgezet onderwijs maar een klein gedeelte van het curriculum met genormeerde toetsen gemeten wordt. Meestal worden alleen het niveau van begrijpend lezen, woordenschat en rekenen met een landelijk genormeerde toets bepaald. Dit betekent dat scholen wel zicht hebben op hun prestaties op een aantal kernvakken, maar dat dit lang niet dekkend is voor het gehele leerstofaanbod van een school.

Samenvattend is het volgende van belang bij het meten van leerwinst:

- de meetinstrumenten zijn dekkend voor datgene wat moet worden gemeten;
- de afdnamecondities bij verschillende leerlingen en scholen zijn vergelijkbaar;
- cohorten leerlingen worden longitudinaal onderzocht;
- de scores op verschillende meetmomenten worden uitgedrukt op dezelfde schaal;
- het meetinstrument is genormeerd.

- *Conclusie: Het is goed mogelijk (voor scholen zelf) om leerwinst inzichtelijk te maken met behulp van data waarover scholen beschikken.*
- *Aanbeveling: Voor niet-cognitieve vaardigheden zouden er meetinstrumenten moeten komen, die voldoen aan de criteria voor het meten van leerwinst.*

7.2 Welke mogelijkheden zijn er om op basis van de informatie die de school zelf heeft een indicatie van de toegevoegde waarde te verkrijgen en mogelijk tot een gestandaardiseerde methode te komen?

Toegevoegde waarde inzichtelijk maken

De toegevoegde waarde van een school is het aandeel in de leerwinst dat aan de school toe te schrijven is. In deze pilot is het niet gelukt om de toegevoegde waarde van een school op basis van de leerwinst inzichtelijk te maken. De leerwinst moet daarvoor gecorrigeerd worden voor leerlingenkenmerken. In hoofdstuk 5 van dit

rapport hebben we laten zien dat kenmerken zoals afkomst uit een armoedeprobleemcumulatiegebied (APCG), leerwondersteuning en indicaties voor een lage afsluiting van het basisonderwijs ('voordeel van de twijfel') samenhangen met de behaalde toetsscores. Zo blijkt bijvoorbeeld dat afkomst uit een APCG-gebied de leerwinst beïnvloedt. Een school met veel leerlingen uit een APCG-gebied kan dus niet zomaar vergeleken worden met een school met weinig van dergelijke leerlingen. Op twee exact dezelfde scholen die alleen verschillen in leerlingenpopulatie is te verwachten dat de school met meer leerlingen afkomstig uit een APCG-gebied een lagere leerwinst zal realiseren. Dit verschil in leerwinst is in zo'n geval niet aan de school toe te schrijven en zegt daarmee weinig over de toegevoegde waarde van de school. Voor scholen is het niet mogelijk om zelf de leerwinst te corrigeren voor leerlingenkenmerken, omdat daarvoor landelijk representatieve gegevens van andere scholen nodig zijn. Daarbij komt dat sommige kenmerken die de resultaten sterk zullen beïnvloeden, zoals het opleidingsniveau van de ouders, niet bekend zijn bij vo-scholen.

Wij hebben geprobeerd de toegevoegde waarde te bepalen uit leerwinstmetingen van de pilotscholen. Daarvoor zijn de aanwezige leerlingenkenmerken verzameld en is in een statistische analyse de leerwinst hiervoor gecorrigeerd om de toegevoegde waarde van de scholen te bepalen. De zo bepaalde verschillen in toegevoegde waarde tussen de pilotscholen waren echter relatief klein en statistisch niet significant. De verschillen tussen scholen bleken niet groter dan de verschillen tussen opeenvolgende cohorten binnen dezelfde school.

Hierbij moet worden opgemerkt dat de aantallen scholen in deze vergelijkingen gering waren. De meest gebruikte toets in deze pilot, de Cito VVO-toets, werd afgenomen op 13 respectievelijk 15 pilotscholen²⁰, die onderling ook nog eens sterk verschilden in het niveau (vmbo, havo, vwo) van het aangeboden onderwijs. Om de mogelijkheden voor het bepalen van toegevoegde waarde te onderzoeken zijn meer gegevens nodig, met name van een groter aantal, landelijk representatieve scholen die allemaal dezelfde toetsen afnemen. Dat was niet de opzet van deze pilot, waarbij scholen juist zelf mochten aangeven met welke meetinstrument ze leerwinst inzichtelijk wilden maken. Het is belangrijk dat scholen dit eigenaarschap over hun eigen instrumenten behouden (zie ook Onderwijsraad, 2014). Scholen zijn namelijk primair zelf verantwoordelijk voor het behoud en de verbetering van de kwaliteit van hun onderwijs. Het is dus de vraag of er verder onderzoek verricht moet worden naar het op deze wijze bepalen van de toegevoegde waarde van scholen. Een effectiever gebruik van leerwinstdata en leerlingenkenmerken is meer voor de hand liggend.

- *Conclusie: Het is vooralsnog niet mogelijk voor scholen om zelf hun toegevoegde waarde te bepalen.*
- *Aanbeveling: Scholen moeten zich richten op hun leerwinst, niet op hun toegevoegde waarde.*

7.3 Welke rol kan het meten van de leerwinst van leerlingen en het mogelijk meten van een toegevoegde waarde spelen bij de verdere ontwikkeling van het opbrengstgericht werken?

Leerwinst en opbrengstgericht werken

Voor het gebruik van leerwinst bij opbrengstgericht werken is de laatste jaren veel aandacht geweest. In verschillende publicaties wordt het belang hiervan benadrukt (School aan Zet (Vermaas, 2013), de VO-raad en de AOC Raad (2014) en de Onderwijsraad (2014)). In het algemeen staat bij opbrengstgericht werken het formuleren en meten van leerdoelen centraal. Het is van groot belang dat deze doelstellingen leerwinst betreffen, en niet alleen eindprestaties. Alleen met leerwinstmetingen kan inzichtelijk worden gemaakt of het

²⁰ 13 in het eerste pilotjaar en 15 in het tweede pilotjaar.

beste uit de leerlingen is gehaald. Dat leerlingen goed scoren bij een eindprestatie betekent niet altijd dat er ook sprake is van een goede leerwinst. Andersom kan bij een lage eindprestatie wel sprake zijn van veel leerwinst. In hoofdstuk 4 is dit inzichtelijk gemaakt door in grafieken de eindprestatie (in dit geval Nederlandse leesvaardigheid) af te zetten tegen de leerwinst. Het is dus belangrijk dat scholen bij opbrengstgericht werken niet alleen sturen op eindprestaties, maar ook op leerwinst. Het verhogen van eindprestaties kan namelijk het eenvoudigst door het instroomniveau te verhogen. Voor het verhogen van leerwinst geldt dit niet. Ook (of misschien juist) bij leerlingen met een relatief laag instroomniveau kan veel leerwinst worden gerealiseerd. Een ander voordeel van leerwinst ten opzichte van eindprestaties is dat voor leerlingen zelf hun vooruitgang inzichtelijk wordt gemaakt. Een positieve leerwinst zal leerlingen meer motiveren dan een tegenvallende eindprestatie. Door inzicht te hebben in de leerwinst kan ook 'tijdens de rit' bijgestuurd worden, wat niet mogelijk is wanneer alleen de eindprestaties worden gemeten.

Het verzamelen en benutten van leerwinstdata

Meerdere toetsystemen bieden de mogelijkheid om leerwinst op schoolniveau inzichtelijk te maken. Uit de pilot blijkt echter dat scholen hier slechts in beperkte mate gebruik van maken. De toetsen worden vooral gebruikt voor de determinatie van individuele leerlingen na het eerste of tweede leerjaar en om vast te stellen welke leerlingen in aanmerking komen voor remedial teaching. Het analyseren van de leerwinst op een hoger niveau zoals de klas, het vak, de docent of de school, komt maar op een deel van de scholen voor. De gebruiksmogelijkheden van de beschikbare data blijven daardoor op veel scholen ten dele onbenut. Bij de verdere ontwikkeling van opbrengstgericht werken is dus niet het verzamelen van gegevens om de leerwinst te bepalen de beperkende factor. Dit kan al (voor een beperkt aantal vakken) en gebeurt al op veel scholen. Het probleem zit in het benutten van de beschikbare gegevens om inzicht te krijgen in de leerwinst op het niveau van een klas, een docent, een vak of de gehele school. Redenen hiervoor zijn dat de scholen niet altijd weten welke rapportage- en analysemogelijkheden hun toets- en administratiesystemen hebben. Maar misschien nog wel belangrijker, scholen vinden het moeilijk om de rapportages te interpreteren en vertalen naar acties voor de dagelijkse praktijk.

In deze pilot zijn de leerwinstgrafieken steeds door Oberon aan de scholen toegelicht. De rapportages zijn in essentie vrij eenvoudig, een vergelijking van de schoolresultaten met een norm. Toch bleek een mondelinge toelichting, in ieder geval bij de eerste ronde schoolrapporten, wel noodzakelijk. Mogelijk onderschatten toetsleveranciers de drempel die scholen moeten nemen om de toetsresultaten te gebruiken voor het maken van rapportages en deze vervolgens te analyseren, te interpreteren en te vertalen naar verbeteracties. De volgende stap voor het gebruik van leerwinst bij het opbrengstgericht werken is voor veel scholen dus niet het creëren van data, maar het beter benutten van de data die al beschikbaar zijn. Het combineren van data over leerwinst met data over achtergrondkenmerken en beginniveau van leerlingen biedt daarbij duidelijk voordelen. Het is daarom in de eerste plaats van belang dat scholen beide soorten data volledig registreren. Uit de pilot blijkt dat lang niet alle scholen data over het beginniveau van leerlingen systematisch bijhouden. Ten tweede is het van belang dat data over achtergrondkenmerken en beginniveau van leerlingen automatisch overgezet kunnen worden tussen het toets- en administratiesysteem waarin deze geregistreerd worden. Volgens de scholen is deze uitwisseling nu vaak nog een belemmerende factor.

Een aantal scholen heeft suggesties gedaan voor het verbeteren van opbrengstgericht werken, bijvoorbeeld scholing van docenten en schoolleiding en leren van elkaar. Meer aanbevelingen zijn te vinden in een recente onderzoekspublicatie van Oberon en KBA (Oomens, e.a., 2015) over het gebruik van gegevens door scholen voor kwaliteitsverbetering. Deze aanbevelingen betreffen onder meer het ondersteunen van scholen bij het realiseren van een onderzoekende en kwaliteitsgerichte cultuur, het verspreiden van goede voorbeelden en (bij)scholing. Ook voor het versterken van opbrengstgericht werken en het optimaler benutten van de reeds

beschikbare data zijn deze aanbevelingen van toepassing. Wat betreft het onderdeel bijscholing kan daarbij gedacht worden aan brede verspreiding van de opgedane ervaringen bij de ondersteuning van de pilotscholen.

- *Conclusie: Naast de eindprestaties is de leerwinst essentieel voor opbrengstgericht werken.*
- *Conclusie: Veel scholen gebruiken de beschikbare data niet optimaal.*

- *Aanbeveling: Help scholen om de vertaalslag te maken van leerwinstdata naar concrete acties.*
- *Aanbeveling: Verbeter de gebruikersvriendelijkheid toets- en administratiesystemen, met name wat betreft de rapportages en de uitwisseling van data tussen deze twee soorten systemen.*
- *Aanbeveling: Houd gegevens over achtergrondkenmerken en uitgangssituatie van leerlingen systematisch bij.*
- *Aanbeveling: Zorg voor bekendheid van 'good practice' scholen zodat scholen van elkaar kunnen leren.*

7.4 Welke ontwikkelingen binnen de pilots en welke producten uit de pilots kunnen worden ingezet bij het verder ontwikkelen van de opbrengstsystematiek en de ontwikkelingen ten aanzien van het (nieuwe) waarderingskader van de inspectie?

Onderscheid voldoende en goede kwaliteitszorg

In de brief 'Transitie in onderwijstoezicht' (Ministerie van OCW, 2014) lichten de Minister en de Staatssecretaris van OCW de plannen voor de verdere ontwikkeling van het toezicht toe. Centraal hierin staat het onderscheid tussen scholen van voldoende en van goede kwaliteit in het funderend onderwijs. In de brief staat: "Een goede school is een school waar leraren onder 1) gunstige schoolcondities, 2) goed onderwijs geven, zodat 3) alle leerlingen optimale opbrengsten realiseren." Het inzichtelijk maken van leerwinst is essentieel om te bepalen of alle leerlingen optimale opbrengsten realiseren. De inspectie kan dit stimuleren door het gebruik van leerwinst te belonen met het oordeel 'goed' op relevante aspecten van de kwaliteitszorg.

Toepassen bij pilots gedifferentieerd toezicht

In het voorjaar van 2015 voert de inspectie in het voortgezet onderwijs pilots gedifferentieerd toezicht uit. Een belangrijk aspect van deze pilots is het beoordelen van indicatoren als voldoende of als goed, bijvoorbeeld of de kwaliteitszorg voldoende of goed is. Bij een aantal van deze pilots wordt gekeken of de mate waarin scholen zelf hun leerwinst inzichtelijk maken, gebruikt kan worden bij het beoordelen van de kwaliteitszorg. Een centrale vraag daarbij is of de vastgestelde leerwinst gevolgen heeft voor de inrichting van het onderwijs. De inspectie wil met deze aanpak het gebruik van leerwinst bij het behoud en de verbetering van de kwaliteit van het onderwijs stimuleren.

Rekening houden met achtergrondkenmerken

Een aantal pilotscholen heeft de wens uitgesproken dat de inspectie bij het beoordelen van hun school meer rekening houdt met de achtergrondkenmerken van hun leerlingen. Om hierover het gesprek aan te gaan met de inspectie is het van groot belang dat de scholen inzicht hebben in de achtergrondkenmerken van hun leerlingen en deze koppelen aan de gerealiseerde leerwinst. Wanneer zij in het bezit zijn van deze gegevens zijn scholen een sterke gesprekspartner voor de inspectie. Het rekening houden met achtergrondkenmerken van leerlingen werkt overigens twee kanten op. Enerzijds de 'uitdagende' leerlingen die misschien wel voldoende leerwinst realiseren. Maar anderzijds ook de 'kansrijke' leerlingen die misschien toch onvoldoende leerwinst realiseren.

Niet-cognitieve opbrengsten

Een aantal pilotscholen heeft de wens uitgesproken dat de inspectie ook naar niet-cognitieve opbrengsten kijkt. De inspectie wil dat zelf ook. In 2015 wordt bijvoorbeeld een themaonderzoek naar burgerschapsvorming uitgevoerd. Sociale en maatschappelijke competenties van leerlingen zijn al onderdeel van het waarderingskader van de inspectie, maar deze blijken in de praktijk moeilijk te beoordelen. Er zijn momenteel nog onvoldoende geschikte genormeerde meetinstrumenten om niet-cognitieve vaardigheden inzichtelijk te maken. Om aan de wens van scholen en van de inspectie te voldoen is het, zoals in paragraaf 7.1 ook al aangegeven, belangrijk dat geschikte meetinstrumenten wel worden ontwikkeld.

Gebruik leerwinst bij het beoordelen van de opbrengsten, verschillen po- en vo-pilot

In januari 2014 is de eindrapportage van de pilot leerwinst en toegevoegde waarde in het primair onderwijs (po) gepubliceerd. Uit de po-pilot blijken meer mogelijkheden om scholen onderling te vergelijken dan uit de vo-pilot. Dit komt doordat bij de po-pilot uitsluitend gebruik werd gemaakt van het Cito volgsysteem primair onderwijs. Dit wordt door de meeste basisscholen gebruikt en is dekkend voor een belangrijk deel van het curriculum van het primair onderwijs (taal en rekenen) en voor alle leerjaren. In het vo is de situatie anders. Lang niet alle scholen nemen landelijk genormeerde toetsen af. Er is een grote verscheidenheid in de landelijk genormeerde toetsen die wel worden afgenomen (onder andere Diataal, Cito, TOA, Route VO en Deviant toetsen). De toetsen die worden gebruikt, dekken maar een beperkt deel van het curriculum in het voortgezet onderwijs (meestal de kernvakken Nederlands, Engels en rekenen/wiskunde) en de toetsen zijn alleen geschikt voor de onderbouw.

Het primair onderwijs heeft dus via het Cito-leerlingvolgsysteem voor de meeste scholen landelijk genormeerde metingen die (grotendeels) dekkend zijn voor het curriculum. Voor het voortgezet onderwijs geldt dit niet. De uitkomsten van de po-pilot wijzen op mogelijkheden om, na verdere ontwikkeling van de systematiek, de toegevoegde waarde van scholen inzichtelijk te maken. De uitkomsten van de vo-pilot wijzen eerder op een gebruik door scholen zelf bij het bewaken en verbeteren van de onderwijskwaliteit, waarbij de inspectie dit meeneemt bij het beoordelen van de kwaliteitszorg.

- *Conclusie: Het is op deze wijze niet mogelijk om de leerwinst van de scholen onderling te vergelijken.*
- *Aanbeveling: Betrek inzicht in leerwinst bij het beoordelen van de kwaliteitszorg van scholen door de inspectie.*

Referenties

- Aitkin, M. & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society*, 149, 1-43.
- Bosker, R. (2012). De toegevoegde waarde van een school: begripsbepaling, meting en causale attributie. In: Dijkstra, A.B. & Janssens, F.J.G. (red) (2012). *Om de kwaliteit van het onderwijs. Kwaliteitsbepaling en kwaliteitsbevordering*. Den Haag: Boom.
- Bosker, R.J., & Luyten, H. (2000). De stabiliteit en consistentie van differentiële schooleffecten. *Tijdschrift voor Onderwijsresearch*, 24(4), 308-321.
- Claassen, A., Driesen, G., Aarntzen, D. & Mulder, L. (2005). *Voorstel voor een indicator voor de toedeling van middelen voor het Leerplusarrangement in het voortgezet onderwijs*. Nijmegen: ITS.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dam, G. ten, Geijssel, F., Reumerman, R., & Ledoux, G. (2010). Burgerschapscompetenties: de ontwikkeling van een meetinstrument. *Pedagogische Studiën*, 87(5) 313-333.
- Dijkstra, A.B. & Janssens, F.J.G. (red) (2012). *Om de kwaliteit van het onderwijs. Kwaliteitsbepaling en kwaliteitsbevordering*. Den Haag: Boom.
- Driessen, G., Mulder, L., & Roeleveld, J. (2012). *Cohortonderzoek COOI⁵⁻¹⁸. Technisch rapport basisonderwijs, tweede meting 2010/11*. Nijmegen: ITS / Amsterdam: Kohnstamm Instituut.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Griffin.
- Guldmond, H. & Bosker, R. (2006). *Onderwijsachterstanden en ontwikkelingen in leerprestaties*. Groningen: GION.
- Hacquebord, H., Stellingwerf, B, Linthorst, R. & Andringa, S. (2005) *Diataal. Verantwoording en normering*. Groningen: Rijksuniversiteit Groningen.
- Inspectie van het Onderwijs (2010). *Opbrengstgericht werken in het basisonderwijs*. Utrecht: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2012). *Projectplan pilot leerwinst/toegevoegde waarde in het voortgezet onderwijs*. Utrecht: Inspectie van het Onderwijs.
- Janssens, F.J.G., Rekers-Mombarg, L. & Lacor, E. (2014). *Leerwinst en toegevoegde waarde in het primair onderwijs. Eindrapportage*. Groningen: GION.
- Ledoux, G., Blok, H., & Boogaard, M.; m.m.v. Krüger, M. (2009). *Opbrengstgericht werken. Over de waarde van meetgestuurd onderwijs*. Amsterdam: SCO-Kohnstamm Instituut.
- McInerney, D.M. & Ali, J. (2006). Multidimensional and hierarchical assessment of school motivation: cross-cultural validation. *Educational Psychology*, 26 (6), 717-734.
- Meijer, J., Ledoux, G., & Elshof, D. (2011). *Gebruikersvriendelijke leerlingvolgsystemen in het primair onderwijs*. Amsterdam: Kohnstamm Instituut.
- Midgley, C., Maehr, M.L., Hruda, L.Z., Anderman, E., Anderman, L., Freeman, K.E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M.J., Nelson, J., Roeser, R., & Urda, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS)*. Ann Arbor, MI: University of Michigan.
- Ministerie van OCW (2011). *Actieplan Basis voor Presteren. Naar een ambitieuze leercultuur voor alle leerlingen*. Den Haag: OCW.
- Ministerie van OCW (2011). *Actieplan Beter Presteren: opbrengstgericht en ambitieus*. Den Haag: OCW.

- Ministerie van OCW (2011). *Actieplan Leraar 2020. Een krachtig beroep*. Den Haag: OCW.
- Ministerie van OCW (2013). *Lerarenagenda 2013-2020: de leraar maakt het verschil*. Den Haag: OCW.
- Ministerie van OCW (2014). *Transitie in onderwijstoezicht*, (kamerstuk 33905 nr.1).
- Ministerie van OCW & VO-raad (2014). *Sectorakkoord VO 2014-2017*.
- Ministerie van OCW & PO-raad (2014). *Bestuursakkoord voor de sector primair onderwijs*.
- Mooij, T., Roeleveld, J., Fettelaar, D. & Ledoux, G. (2012). Kwaliteitsbeoordeling van scholen primair onderwijs: Het correctiemodel van de inspectie vergeleken met alternatieve modellen. *Pedagogische Studiën*, 89, pp. 272-287.
- OECD (2008). *Measuring Improvements in Learning Outcomes. Best Practices to Assess the Value-Added of Schools*. Parijs: OECD
- Onderwijsraad (2003). *Wat scholen toevoegen. Instrumenten voor de bepaling van de toegevoegde waarde van het basisonderwijs*. Den Haag: Onderwijsraad.
- Onderwijsraad (2013). *Vooruitgang boeken met achterstandsmiddelen*. Den Haag: Onderwijsraad.
- Onderwijsraad (2014). *Toegevoegde waarde: een instrument voor onderwijsverbetering – niet voor beoordeling*. Den Haag: Onderwijsraad.
- Oomens, M., Aarsen van, E. & Hulsen, M. (2012). *Gebruik van en ervaringen met leerlingvolgsystemen in het voortgezet onderwijs*. Utrecht: Oberon.
- Oomens, M., Buynsters, M., Donker, E., Geldhof, T., Boer, P. den, Frietman, J., Verhaegh, T. & Ehren, M. (2015). *Informatiegebruik voor kwaliteitsverbetering*. Utrecht/Nijmegen: Oberon/KBA.
- Rasbash, J., Charlton, C., Browne, W.J., Healy, M. & Cameron, B. (2005). *MLwiN Version 2.02*. Centre for Multilevel Modelling, University of Bristol.
- Raudenbusch, S. W. (2004). *Schooling, statistics, and poverty: Can we measure school improvement?* Princeton, NJ: Educational Testing Service. Opgehaald op 13 maart 2012, van www.ets.org/Media/Education_Topics/pdf/angoff9.pdf.
- Raudenbush, S. & Bryk, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.
- Rekers-Mombarg, L.T.M., Timmermans, A.C., Bosker, R.J. (2014). *Vernieuwing van indicatoren voor leeropbrengsten in het voortgezet onderwijs; een dieptestudie*. Groningen: GION.
- Roeleveld, J. (1994) *Verschillen tussen scholen. Kenmerken, effectiviteit en stabiliteit van onderwijsinstellingen in Nederland*. Proefschrift Universiteit van Amsterdam.
- Roeleveld, J. (2003). *Herkomstkenmerken en begintoeets. Secundaire analyse op het PRIMA-cohortonderzoek*. Studie voor de Onderwijsraad.
- Roeleveld, J. (2011). De cognitieve ontwikkeling van doelgroepoerlingen gedurende het basisonderwijs. In: Roeleveld, J., Driessen, G., Ledoux, G., Cuppen, J. & Meijer, J. (2011). *Doelgroepoerlingen in het basisonderwijs. Historische ontwikkeling en actuele situatie*. pp. 61-88 Amsterdam/Nijmegen: Kohnstamm Instituut/ITS.
- Roeleveld, J., Veen, I. van der & Ledoux, G. (2008). *Verkenning leerwinst als indicator voor onderwijskwaliteit*. Onderzoek voor het Ministerie van Onderwijs, Cultuur en Wetenschappen. Amsterdam: SCO-Kohnstamm Instituut.
- Roeleveld, J., Mooij, T., Fettelaar, D. & G. Ledoux (2011). *Correctiefactoren bij opbrengstmaten in het primair onderwijs*. Amsterdam / Nijmegen: Kohnstamm Instituut / ITS.
- SCP/CBS(2007). *Armoedemonitor 2007*. Sociaal en Cultureel Planbureau, Den Haag.

- Seegers, G., Putten, C.M. van, & Brabander, C.J. de (2002). Goal orientation, perceived task outcome and task demands in mathematics tasks: Effects on students' attitude in actual task settings. *British Journal of Educational Psychology*, 72(3), 365-384.
- Timmermans, A. (2012). *Value added in educational accountability: Possible, fair and useful?* Groningen, Proefschrift RUG.
- Veen, I. van der & Peetsma, T. (2009). The development in self-regulated learning behavior of first-year students in the lowest level of secondary school in the Netherlands. *Learning and Individual differences*, 19(1), 34 - 46.
- Vermaas, J. (2013). *Opbrengstgericht werken en het vakmanschap van de leraar. OGW in het VO*. Den Haag: School aan Zet.
- VO-raad en AOC-raad (2014). *De toolkit OGW, samen werken aan onderwijskwaliteit*.
- Vreeburg, B. (2008). *Absoluut versus relatief beoordelen van de opbrengsten van scholen*. Paper Onderwijs Researchdagen 2008.
- Wijnstra, J., Ouwens, M., & Béguin, A. (2003). *De toegevoegde waarde van de basisschool. Verkenning van de mogelijkheden de schoolspecifieke bijdrage aan de onderwijsopbrengst in kaart te brengen met behulp van het Cito Leerlingvolgsysteem en de Eindtoets Basisonderwijs*. Arnhem: Citogroep.
- Zijsling, D., Keuning, J., Naayer, H. & Kuyper, H. (2012). *Cohortonderzoek5-18. Technisch rapport meting VO-3 in 2011*. Groningen / Arnhem: GION / Cito.

Websites

<http://haal-eruit-wat-erin-zit.nl> Handreiking voor datagebruik en -analyse bij opbrengstgericht werken van School aan Zet.

<http://www.onderwijsinspectie.nl/onderwijs/Voortgezet+onderwijs/leerwinst-en-toegevoegde-waarde>
Informatie over de pilot.

Bijlage 1 Projectgroep en leveranciers

Samenstelling projectgroep

Alex Coenen, Inspectie van het onderwijs

Ankie Bosman, Ministerie van OCW

Inge Drewes, Inspectie van het onderwijs

Jaap Roeleveld, Kohnstamm instituut

Job Goossens, Inspectie van het onderwijs

Margot Oomens, Oberon

Leveranciers toetsystemen waarmee gesprekken zijn gevoerd

A-VISION, (Route VO)

Bureau ICE (TOA)

Cito

Diataal

Uitgeverij Deviant

Bijlage 2 Voorbeelden 'Leerlingeninstroom' uit de schoolrapporten

Voorbeeldtabel 1 Achtergrondkenmerken leerlingen in leerjaar 1 en 3; landelijk en op uw school

landelijk eerste leerjaar	% meisjes	% APCG	% lwoo	% vertraagd bij instroom
vmbo-basis en kader	45	26	49	46
vmbo-breed (incidenteel + havo)	49	18	38	39
vmbo-(g)t	52	17	5	26
vmbo-breed/havo/vwo	50	10	12	20
vmbo-(g)t/havo/vwo	50	12	1	14
havo/vwo	49	13	0	9
vwo	46	14	0	4
uw pilot afdeling				
vmbo-basis leerjaar 1 2013/2014	54	0	78	37
vmbo-kader leerjaar 1 2013/2014	63	0	38	26
vmbo-tl/gl leerjaar 1 2013/2014	65	0	11	16
aantal leerlingen met gegevens				
vmbo-basis leerjaar 1 2013/2014	75	75	75	75
vmbo-kader leerjaar 1 2013/2014	103	103	103	103
vmbo-tl/gl leerjaar 1 2013/2014	75	75	75	75
landelijk derde leerjaar	% meisjes	% APCG	% lwoo	
vmbo-basis	44	22	64	
vmbo-kader	47	17	26	
vmbo-(g)t	49	13	5	
havo	51	10	0	
vwo	53	9	0	
uw pilot afdeling				
vmbo-basis leerjaar 3 2013/2014	58	2	85	
vmbo-kader leerjaar 3 2013/2014	63	0	46	
vmbo-tl/gl leerjaar 3 2013/2014	58	2	17	
aantal leerlingen met gegevens				
vmbo-basis leerjaar 3 2013/2014	59	59	59	
vmbo-kader leerjaar 3 2013/2014	108	108	108	
vmbo-tl/gl leerjaar 3 2013/2014	83	83	83	

Voorbeeldtabel 2 Kenmerken afsluiting basisonderwijs bij leerlingen in leerjaar 1 en 3; landelijk en op uw school

landelijk eerste leerjaar	% advies was hoger	% advies was lager	% Cito-EB was hoger	% Cito-EB was lager	% NIO was hoger	% NIO was lager
vmbo-basis en kader	17	2	23	0		
vmbo-(g)t	4	5	15	25		
havo/vwo	0	1	0	7		
vwo	0	1	0	20		
uw pilot afdeling						
brugklas havo/vwo leerjaar 1 2013/2014	0	0	0	13	0	34
aantal leerlingen met gegevens						
brugklas havo/vwo leerjaar 1 2013/2014	103		95		88	
landelijk derde leerjaar	% advies was hoger	% advies was lager	% Cito-EB was hoger	% Cito-EB was lager	% NIO was hoger	% NIO was lager
vmbo-basis	31	2	31	-		
vmbo-kader	32	14	38	32		
vmbo-(g)t	20	9	28	19		
havo	9	10	13	25		
vwo	-	14	-	39		
uw pilot afdeling						
havo leerjaar 3 2013/2014	9	0	23	18	-	-
vwo leerjaar 3 2013/2014	0	10	0	33	-	-
aantal leerlingen met gegevens						
havo leerjaar 3 2013/2014	46		40		0	
vwo leerjaar 3 2013/2014	31		27		0	

Voorbeeldtabel 3 Gemiddelde scores op Cito Toets 0 naar leerlingkenmerken; vmbo-kader

	Ned. lees- vaardigheid	Ned. woord- schat	Eng. lees- vaardigheid	rekenen/ wiskunde	aantal
landelijke norm basis	200	204	197	191	
landelijke norm kader	206	211	205	198	
landelijke norm gl/tl	213	219	214	212	
uw afdeling kader totaal	204	198	214	199	109
jongens	201	196	215	207	50
meisjes	207	200	212	192	59
onvertraagd bij instroom	208	207	214	200	62
vertraagd bij instroom	199	187	213	198	47
zorgindicatie(lwoo)	198	189	206	195	29
geen indicatie	206	202	216	201	80
APCG-gebied	205	202	215	199	95
overig	200	175	203	198	14
advies lager dan afdeling	193	179	196	202	2
advies gelijk	210	211	216	202	45
advies hoger dan afdeling	205	200	218	201	46
lage Cito score	209	212	215	197	20
gem. Cito (voor afdeling)	210	209	214	201	35
hoge Cito	213	222	221	213	20
lage NIO score	205	200	216	197	12
gem. NIO (voor afdeling)	214	223	214	208	11
hoge NIO	192	195	176	215	2

Voorbeeldtabel 4 Beginscores Begrijpend lezen (Diataal 1A uit 2009/2010) bij afdeling vmbo-tl vergeleken met het CE-cijfer Nederlands (eindexamen 2013/2014).

Kwartielen Begrijpend Lezen	CE-cijfer			aantal
	% onvoldoende	% voldoende	% goed	
0-25%	50%	42%	8%	12
25-50%	25%	67%	8%	12
50-75%	0%	58%	42%	12
75-100%	31%	46%	23%	13
allen	27	53	20	49
landelijk vmbo-tl	20	35	45	



Ministerie van Onderwijs, Cultuur en
Wetenschap



Inspectie van het Onderwijs
Ministerie van Onderwijs, Cultuur en
Wetenschap

Oberon
onderzoek | advies

KOHNSTAMM
INSTITUUT
Kohnstamm